

Article

An Ensemble-Based Protocol for the Computational Prediction of Helix-Helix Interactions in G Protein-Coupled Receptors using Coarse-Grained Molecular Dynamics

Nojood A. Altwaijry, Michael Baron, David W. Wright, Peter Vivian Coveney, and Andrea Townsend-Nicholson

J. Chem. Theory Comput., **Just Accepted Manuscript** • DOI: 10.1021/acs.jctc.6b01246 • Publication Date (Web): 06 Apr 2017

Downloaded from <http://pubs.acs.org> on April 12, 2017

Just Accepted

“Just Accepted” manuscripts have been peer-reviewed and accepted for publication. They are posted online prior to technical editing, formatting for publication and author proofing. The American Chemical Society provides “Just Accepted” as a free service to the research community to expedite the dissemination of scientific material as soon as possible after acceptance. “Just Accepted” manuscripts appear in full in PDF format accompanied by an HTML abstract. “Just Accepted” manuscripts have been fully peer reviewed, but should not be considered the official version of record. They are accessible to all readers and citable by the Digital Object Identifier (DOI®). “Just Accepted” is an optional service offered to authors. Therefore, the “Just Accepted” Web site may not include all articles that will be published in the journal. After a manuscript is technically edited and formatted, it will be removed from the “Just Accepted” Web site and published as an ASAP article. Note that technical editing may introduce minor changes to the manuscript text and/or graphics which could affect content, and all legal disclaimers and ethical guidelines that apply to the journal pertain. ACS cannot be held responsible for errors or consequences arising from the use of information contained in these “Just Accepted” manuscripts.



ACS Publications

An Ensemble-Based Protocol for the Computational Prediction of Helix-Helix Interactions in G Protein-Coupled Receptors using Coarse-Grained Molecular Dynamics

Nojood A. Altwaijry^{1,3}, Michael Baron¹, David W. Wright², Peter V. Coveney² and Andrea Townsend-Nicholson^{1,}*

AUTHOR ADDRESS

¹ Institute of Structural & Molecular Biology, Research Department of Structural & Molecular Biology, Division of Biosciences, University College London, London, WC1E 6BT, United Kingdom

² Centre for Computational Science, Department of Chemistry, University College London, London WC1H 0AJ, United Kingdom

³ King Saud University, Riyadh, Kingdom of Saudi Arabia

KEYWORDS Computational biology, computational chemistry, ensemble-based molecular dynamics, coarse-grained molecular dynamics, CG-MD, G protein-coupled receptor, GPCR, adenosine receptor, A_{2A} receptor, rhodopsin, helix-helix interactions

ABSTRACT

The accurate identification of the specific points of interaction between G protein-coupled receptor (GPCR) oligomers is essential for the design of receptor ligands targeting oligomeric receptor targets. A coarse-grained molecular dynamics computer simulation approach would provide a compelling means of identifying these specific protein-protein interactions and could be applied both for known oligomers of interest and as a high-throughput screen to identify novel oligomeric targets. However, to be effective, this *in silico* modeling must provide accurate, precise, and reproducible information. This has been achieved recently in numerous biological systems using an ensemble-based all-atom molecular dynamics approach. In this study, we describe an equivalent methodology for ensemble-based coarse-grained simulations. We report the performance of this method when applied to four different GPCRs known to oligomerize, using error analysis to determine the ensemble size and individual replica simulation time required. Our measurements of distance between residues shown to be involved in oligomerization of the fifth transmembrane domain from the adenosine A_{2A} receptor are in very good agreement with the existing biophysical data and provide information about the nature of the contact interface that cannot be determined experimentally. Calculations of distance between rhodopsin, CXCR4 and β_1 AR transmembrane domains reported to form contact points in homodimers correlate well with the corresponding measurements obtained from the structural

data, providing an ability to predict contact interfaces computationally. Interestingly, error analysis enables identification of non-interacting regions. Our results confirm that GPCR interactions can be reliably predicted using this novel methodology.

1. INTRODUCTION

We need to understand how proteins behave in order to manipulate them successfully. The means by which to achieve accurate, precise, and reproducible predictions of the key properties of therapeutically relevant proteins is a fundamental question in computational biology. Molecular dynamics (MD) simulations have been used to study complex biomolecular systems, but it is not possible to define how a system behaves from a single trajectory; single trajectory systems behave as Gaussian random processes, making the attainment of accurate predictions from a single run not a realistic proposition. Accurate predictions that correlate well with experimental data have been achieved with the use of multiple, short MD simulations, to enhance the sampling of conformational space and, hence, the convergence of observable properties¹⁻⁷. These ensemble-based fully atomistic MD studies have primarily focused on ligand-protein binding free energies, where there exists a wealth of experimental data with which to compare computational findings. In this paper, we take our first steps to assess reliability and reproducibility of analogous CG-MD simulations. For this work, we have elected to examine the molecular nature of protein-protein interactions between G protein-coupled receptors (GPCRs). This is a biological system with which we are familiar experimentally⁸⁻¹⁴.

GPCRs are a particularly well-studied family of membrane proteins. Not only are they a large and important group of signaling proteins, they are also the targets for about 40% of all

therapeutic compounds in clinical use. Although over 800 human proteins are classified as GPCRs, drugs have been developed against fewer than 10% of these targets¹⁵. Thus there is huge potential to expand the number of targets for which new therapies can be designed. Novel therapeutic design is also important if one of the goals of personalized medicine, to develop new drugs for patient-specific variations of GPCRs, is to be achieved. Inclusion of functional GPCR homomers and heteromers in drug discovery programs also provides a means of expanding the range of novel targets for the development of therapeutic agents¹⁶. Originally believed to function as monomeric proteins, many functional GPCR oligomers have now been identified. Early examples include the obligate heteromeric assembly of GABA_BR1 and GABA_BR2 required to form a functional GABA_B receptor¹⁷ and heterodimerization of the delta and kappa opioid receptor subtypes to form an opioid receptor with the κ_2 receptor subtype pharmacology¹⁸. The archetypal class A GPCR, rhodopsin, forms structural dimers organized in paracrystalline arrays in membranes¹⁹ and in the model crystal structure of this GPCR (1N3M)²⁰. To be able to design cost-effective “designer” drugs for individuals that target receptor oligomers, it will be necessary to develop a powerful and sophisticated computational method of understanding the interactions involved in the formation of GPCR oligomers.

Biological methods for studying GPCR oligomers in native cells and tissues or in recombinant mammalian expression systems include co-immunoprecipitation, western blot analyses, cross-linking studies, yeast two hybrid experiments, bi-molecular fluorescence complementation via GFP reconstitution (BiFC), energy transfer-based methods (FRET and BRET), functional cross-talk and activation by dimeric/bivalent ligands²¹. Unfortunately, these methods frequently allow for alternative interpretations of the results and, therefore, do not provide unequivocal answers about multimerization occurrence between candidate pairs of GPCRs nor do they yield specific

1
2
3 details of the interface(s) between interacting GPCRs. Structural methods such as x-ray
4
5 crystallography and atomic force microscopy could provide some of this information, but only
6
7 three Class A GPCR dimer structures have been solved²²⁻²⁴ and tend to describe “contact areas”
8
9 rather than specific molecular interfaces. In order to develop an accurate computational model
10
11 for analyzing GPCR interfaces, it is essential to have good experimental data with which to
12
13 validate the model. Such data are made available for the A_{2A} adenosine receptor subtype, which
14
15 has been shown to participate in the formation of both heteromeric²⁵ and homomeric GPCRs²⁶.
16
17 The identification of homomeric A_{2A} receptors provided an opportunity to identify the
18
19 transmembrane domain (TM5) involved in the self-association, by far-UV CD spectroscopy and
20
21 SDS-PAGE using synthetic peptides corresponding to the different transmembrane domains²⁷. A
22
23 subsequent study²⁸ mapped TM helix interactions in the A_{2A} receptor for 31 different peptide
24
25 pairs. We have previously worked with the A_{2A} receptor gene²⁹ and are interested in identifying
26
27 patient-specific variations within this and related nucleoside and nucleotide receptor subtypes.
28
29
30
31
32
33

34 There have been many computational studies of GPCR interactions (see Table 1). The
35
36 methodologies for modeling these have, in general, adopted one of two approaches: (i) molecular
37
38 dynamics simulations using models based on homology with the nearest related GPCR for which
39
40 structural data exist or (ii) docking^{30,31}. Initial GPCR MD studies were performed using
41
42 CHARMM and AMBER which were subsequently integrated into NAMD^{32,33} and
43
44 GROMACS^{34,35}. Although there is no established standard protocol for MD simulations of
45
46 GPCRs, a number have used GROMACS with the Martini force field³⁶⁻⁴¹, which is designed
47
48 specifically for lipids and membranes and allows the lipid composition most suited to the
49
50 receptor in question to be incorporated into the simulation. The more recent of the GPCR dimer
51
52 modeling studies have been conducted using coarse-grained simulations, which take less
53
54
55
56
57
58
59
60

compute time and, therefore, provide an opportunity to perform a substantial number of replicas for each set of simulation conditions.

When we began our studies, approximately 30 computational GPCR dimer models had been published. Of these, 2 are Monte Carlo-derived, 15 are based on docking and 9 have been generated using atomistic MD simulations. The rest are CG-MD models. Historically, docking was the earliest method to be employed and has been used regularly; its current use is widespread. Alternative methods of modeling began with Monte Carlo methods, moving to fully atomistic MD, with a subsequent shift to CG-MD, which is the predominant MD method currently in use for GPCRs. CG-MD is popular as it is cheaper, faster and has been shown, when CG models are subsequently converted to atomistic representations, to produce similar results to models generated by atomistic MD^{38,42-43}. CG-MD simulations have also been used to study TM helix-helix dimers for non-GPCR types of cell surface receptors such as Glyphorin A and ErbB dimers^{44,45}.

GPCRs exhibit thermodynamic equilibrium states and, therefore, are “mixing” in the language of ergodic dynamical systems theory⁵. Neighboring trajectories diverge exponentially rapidly and only probabilistic descriptions are meaningful. For these intrinsic reasons, collections of trajectories differing only in their initial conditions, known as ensembles, are the best means of studying the properties of such systems. Each individual system in the ensemble is referred to as a replica. As an additional benefit, performing such ensemble-based molecular dynamics simulations provides close control of errors and uncertainties in predictions. In this paper, we present the development of a robust and rapid method of this kind for identifying helix-helix interactions in GPCRs.

2. METHODS

Here, we aim to develop a consistent, rapid, reproducible CG-MD methodology for the study of interacting helices. This method involves placing two GPCR transmembrane helices (a simulation set) in a membrane and running simulations hoping to identify interactions between the helices. In these simulations, we will be using distance as a means of identifying two different types of interactions: interactions between helices and interactions between amino acid residues on each helix. For the successful identification of both types of interaction, it is necessary to specify the number of replicas (independent simulations identical other than for the initial velocity seeds assigned to the particles) and the run time needed to achieve converged results and see how well they reproduce experimental results. The number of replicas must be sufficient to achieve a reproducible result as evidenced by a sufficiently small error estimate.

We will use the terms “stable dimer” and “dimerization” to refer to interactions between helices. A 10Å truncation cutoff (backbone to backbone) has been set for dimerization, as it has been shown experimentally that a unique FRET signal is generated when two labeled peptides are located within 10Å of each other and form an excited stated dimer⁴⁶. The term “specific interactions” will be used to refer to interactions between amino acid side chains on the dimerized helices. Specific interactions will be identified from contact matrices (heat maps). Although a 12Å truncation cutoff had previously been used to analyze these⁴⁷, we will set our interaction cutoff to 10Å, since the existence of hydrogen bond ($C_{\alpha}-H\cdots O$) contacts as a function of the inter-helical axial distance is between 6 and 12Å. Side chain to side chain distances consistent with this are used to identify specific interactions, with distances of 5-7Å reflecting stronger interactions.

From Table 1, it can be seen that the longest total simulation time for atomistic MD is 0.1 μ s and for CG simulations is 200 μ s. The formation of a long-lasting helix dimer was identified within a few hundred nanoseconds in CG-MD studies of Glucocorticoid Receptor A, a non-GPCR model for studying TM membrane protein structure²³. The number of replicas performed in these different studies varies tremendously, but is never greater than 10. Excellent agreement has been obtained between computed binding free energies and experimental data when ensembles of 50 replicas are used¹. We therefore selected 500 ns for the run time and 50 replicas as starting parameters for these studies. These calculations were run on Legion and Grace, two high-performance Research Computing cluster at UCL (details of the machines used can be found at https://wiki.rc.ucl.ac.uk/wiki/RC_Systems#Legion_technical_specs and https://wiki.rc.ucl.ac.uk/wiki/RC_Systems#Grace_technical_specs). Our preliminary tests showed that CG-simulations (1 ensemble) run on Legion of 500 ns completed within approximately 150 hours. CG-simulations (1 ensemble) run on Grace of 500ns completed within approximately 72 hours.

2.1 CG Simulations

All CG-MD simulations were performed in GROMACS (version 4.6.4) (www.gromacs.org). The temperature was equilibrated for all three groups: protein, lipid bilayer, and solvent (water) with ions to remove center of mass motion relative to the bilayer and protein. The thermalization run was carried out for 100 ps. The simulations were then run at 310 K (the human physiological temperature), which is below the phase transition temperature of pure DPPC (315 K). The system output of the temperature was evaluated to make sure that it stabilized at the required temperature (310K) before continuing until pressure equilibration was attained. An ensemble of

50 replicas for each simulation box (see Tables 2 and 4) was performed. Each simulation was run for 500 ns. CG-atom velocities were drawn from a Maxwell-Boltzmann distribution at $T=310$ K, but all other variables were kept constant; standard deviation was used to compare differences in mean distance outputs. Each simulation was run independently, with the initial configurations differing only by the starting velocity; they were performed under within the NPT ensemble (i.e. constant temperature, pressure, and particle number) using the Martini 2.2 force field⁴⁸. The temperatures of the protein and the lipid were coupled using the velocity-rescaling (modified Berendsen) thermostat at 310 K (human physiological body temperature), with a coupling constant of $T_t = 1$ ps. The system pressure was semi-isotropic using the Berendsen algorithm at 1 bar, with a coupling constant of $T_p = 1$ ps and a compressibility of 1×10^{-4} bar⁻¹. An integration time step of 30 fs was chosen and the coordinates were saved every 10000 subsequent steps for further analysis. The electrostatic interactions were shifted to zero between 0 to 1.2 nm. The Lennard-Jones (LJ) potential was shifted to zero between 0.9 to 1.2 nm to reduce the cutoff noise. The neighbor list for pairwise non-bonded interactions was determined by the Verlet cutoff scheme at 1.4 nm and updated every 10 steps.

2.2 Dimer Analysis

Interhelix distance matrices were calculated for the helix-helix dimer formation and contact maps used to identify specific interactions between residues were generated using the GROMACS tool `g_mdmat`. The individual helix-helix contacts from each replica were examined by calculating the resulting inter-helix distance matrices from the initial simulation starting distance of 4 nm (40 Å) to assess the reproducibility, number of replicas and run time needed to achieve convergence through a locally written code. In those runs where dimerization was

observed, the trajectories were combined and examined in greater depth by calculating the averaged inter-helix distance matrices with three different truncation distances, 15Å, 12Å and 10Å, to determine the dimerization properties between the helices. A cutoff distance of 10Å was then applied to identify residues involved in helix-helix interactions. To investigate the influence of the number of replica simulations on the reliability of our results, we calculated mean interaction distances for ensembles of varying size. For evaluations of run length, mean distance output for the entire ensemble was calculated at 100 ns increments.

Representative atomistic structures of the different CG dimers were generated through use of the “backward” Python script⁴⁹ and the g_cluster tool in GROMACS using the gromos algorithm at a cutoff of 2.5 nm⁵⁰. Visualization was performed using VMD⁵¹. Approximate distances between the atomistic residues in interacting helices were measured using Jmol (www.jmol.org). Amino acid positions have been described using amino acid number in conjunction with the Ballasteros-Weinstein nomenclature⁵² (in superscript). Pairwise combinations used in the analyses were obtained from a matrix of the number of amino acid residues in helix 1 multiplied by the number of amino acid residues in helix 2. In the A_{2A} receptor, there are 729 possible pair combinations between the two 27 residue long TM5 helices, for example, combination 552 specifies the combination of residue 23 (helix 1) with residue 24 (helix 2), representing the V196^{5.57}-Y197^{5.58} interaction.

2.3 Construction of A_{2A}R TM Helices and Preparation of the Simulation Box

Initial simulations were performed using TM5 of the human A_{2A} adenosine receptor, which has been shown experimentally to form a homodimer²⁸. TM2 of the A_{2A} receptor was used as a negative control, as it was unable to form a homodimer under the same experimental conditions.

M193^{5,54}, identified experimentally as being involved in the helical interface and located within a PXXXM motif, and M177^{5,38}, which we identified as residing in a previously unidentified upstream PXXXM motif, were mutated *in silico* (M177^{5,38}A, M193^{5,54}A, and M193^{5,54}I), to permit simulation of the experimental condition in which M193^{5,54} had been mutated and the biological properties of the mutated protein compared with wild type.

The five A_{2A} TM helices shown in Table 2 were generated using MODELLER 9.12 following the procedure detailed^{53,54}, using the crystal structure of the A_{2A} receptor (PDB accession number 3EML; GI: 209447557)⁵⁵. The atomistic helices were subsequently converted into CG models using the “martinize” Python script⁴³. A simulation box of dimensions 8 nm x 8 nm x 8 nm was constructed containing two wild-type TM5 helices (Figure 1). The helices were placed 4 nm apart and aligned in a parallel orientation mimicking the natural positioning of the helix in the membrane (see Figure 1a), with their long axes parallel to the z-axis of the box (see Figure 1b). The TM helices were separated by 4 nm at the beginning of the simulation to rule out any initial inter-helix interactions. Water and lipids were then added. Approximately 190 molecules of the 1, 2-dipalmitoyl-sn-glycero-3-phosphocholine (DPPC) lipid bilayer and additional water molecules (~2660-2690) were added in coarse-grained form in a 3-dimensional cuboid box with periodic boundary conditions using the “insane” Python script⁵⁶. To neutralize the net charge on the protein, water molecules were replaced by counterions (either Na⁺ or Cl⁻, as appropriate, depending on the amino acid composition of the helices).

3. RESULTS

Our aim in the following sections is to investigate the computational parameters required to obtain converged results, to identify whether these results match the experimentally-obtained

data for the self-association of the TM5 helices of the A_{2A} adenosine receptor and, if they do, to further validate these parameters using structural biology data from Class A GPCRs that have been experimentally shown to form a dimeric biological unit.

3.1 Internal Sampling, Convergence, and Reproducibility

In our simulations, the TM helices were observed to diffuse freely in the lipid bilayer. The kernel density estimation of the mean distance between the two wild type TM5 helices at t=0 and at increments of 100ns up to completion of the simulation at 500 ns across the 50 replica ensemble is shown in Figure 2. At t=0, the two helices are at their starting positions 40Å (4.0 nm) apart. At t=500 ns, the mean distance between the helices has adopted a normal distribution with a mean distance of ~16Å between them. The intermediate time points show the redistribution of the distance from the starting point at t=0 to the final mean distance between the helices at 500 ns. A graphical representation of the number and timing of interactions observed in each replica within the ensemble of 50 replicas for the wild-type TM5-TM5 simulation is shown in Figure 3. Four of the 50 replicas showed no contact between the helices, which gives rise to the small peak at 40Å in Figure 2. Three of the replicas began to show contact towards the end of the run; this corresponds to the smaller peak seen at 30Å in Figure 2.

3.1.1 Optimal Replica Number Required

Five different ensembles, one for each A_{2A} receptor simulation set, were run independently in CG-MD simulations for the total run time of 500 ns. These data, which included both wild type and mutated helix sequences, were used to investigate whether variations in the optimal replica number required would occur between different simulation sets. This information was used to

identify the minimum replica number required to achieve convergence for any given simulation set.

Figure 4a reveals that there is no statistically significant difference in the mean distance as a function of replica number. However, a decrease in the error of the mean is observed with increasing ensemble size. From Figure 4b, it can be seen that the rate of decrease in the error slows after approximately 15 replicas are included in the ensemble. For each of the five sets, larger ensembles provide less variation in the error of the mean and an ensemble of 30 replicas represents a good compromise between computational effort and minimization of the error in the mean distance calculated. We conclude that an ensemble of 30 replicas is sufficient to achieve convergence.

3.1.2 Minimum Run Length Required

The effect of run time on the average distance between the helices was examined by calculating the mean distance and the standard deviation within the 50 replicas for simulations of varying duration (0, 100, 200, 300, 400 and 500 ns). Figure 5a shows a significant effect of run length on both mean distance and standard deviation, confirming the results of Figure 3 and reflecting the time required for interactions to take place. For four of the five simulation sets, the standard deviation increases as a function of time, with the rate of increase slowing as the run length becomes longer. In contrast, no change in the standard deviation over time is seen in the TM2-TM2 set (Figure 5b). Interestingly, TM2 homodimers could not be detected experimentally⁵⁷. The absence of an increase in error in the mean distance as a function of time may serve as an indicator of an absence of interaction between two helices within an ensemble.

We conclude that an ensemble run for a simulation time of 300 ns is sufficient to achieve convergence.

3.2 Interacting Interfaces

The final mean distance between the two helices in the ensemble of 50 replicas was used to identify the specific interactions between the A_{2A} homodimers, for each simulation set tested. Following application of the 10Å cutoff, 26% of the ensemble formed stable dimers in the wild-type TM5-TM5 simulations. In the mutated TM5-M177^{5.38}A and in the TM5-M193^{5.54}I simulation sets, 16% of the ensemble formed stable dimers while in TM5-M193^{5.54}A, dimers were detected in 28% of the ensemble. For all four of the TM5 simulation sets, the detected interactions took place at the same position within the helices, indicating that a defined orientation is needed to establish a specific interaction. In the negative control (the TM2-TM2 simulation set), 24% of the ensemble resulted in the formation of stable dimers but there were no specific interactions identified between residues. For all simulations, we combined the trajectories of those pairwise combinations in which dimerization was identified after the cutoff 10Å had been applied and compared the results with heat maps of interactions observed at 12Å and 15Å (see Figure 6). The location of the contact interface was then mapped by comparison with the crystal structure of A_{2A}R (3EML).

3.2.1. Identification of Contact Interface for the Wild-type TM5-TM5 Homodimer

Figure 6 shows the average interhelical contact distance between the two wild-type TM5-TM5 helices (Figure 6a-c) or between the negative control TM2-TM2 helices (Figure 6d-f). The proximity of the wild-type helices is best visualized at 15Å (Figure 6a and c). The interacting

residues in the wild-type TM5-TM5 simulation are found in the bottom third of the C terminal end of TM5. From the averaged interhelix contact matrices, the specific interactions were found to be within the experimentally identified M193^{5,54}xxVY197^{5,58} motif at an interhelical distance of $\sim 8\text{-}9\text{\AA}$. The methionine at position 193^{5,54} of helix 1 interacts with the methionine at the same position on helix 2, reinforcing the suggestion²⁷ of its importance in the formation of the TM5 homodimer. From Figure 6d, it can be seen that the distance between TM2-TM2 is close enough to form potential specific interactions, however, none were detected in the combined trajectories for this negative control. Results obtained at the 15 \AA cutoff (Figure 6f) were random and non-specific, supporting the selection of a minimum cutoff distance of 12 \AA . It should also be noted that there was no increase in the standard deviation over time for the TM2-TM2 simulation (Figure 5b), whereas there was an increase in this variable for all simulation sets in which specific interactions occurred, indicating that the change in error over time may be a useful indicator of helix-helix interactions. The frequency of specific interactions identified in the wild-type TM5-TM5 ensemble was determined by calculating the mean distance for each frame of every replica individually. Table 3 shows that the five most prominently occurring interactions were between M193^{5,54}-M193^{5,54}, V196^{5,57}-Y197^{5,58}, Y197^{5,58}-R199^{5,60}, R199^{5,60}-R199^{5,60} and R199^{5,60}-I200^{5,61}.

These findings are consistent with the experimental results²⁷ identifying that the interaction between two wild-type A_{2A} TM5 peptide sequences involved amino acid residue M193^{5,54}. Our findings are also consistent with experimental data showing the formation of A_{2A} receptor homodimers using bioluminescence resonance energy transfer (BRET)²⁶ and bimolecular fluorescence complementation (BiFC)⁵⁸. The presence of specific interactions between TM2 helices was experimentally investigated and none were detected^{27, 57}. Our CG-MD simulations

produced the same results as experimentally obtained findings, with the formation of wild-type TM5-TM5 dimers involving the M193^{5.54} residue and no specific interaction detected between TM2-TM2 helices *in silico*.

3.2.2 Mutated TM5 Interacting Interfaces

The identification of the presence of M193^{5.54} in the contact interface suggested that this residue may play a significant role in how the two TM5 helices interact. To investigate this possibility, we introduced sets of ensemble simulations that included mutated helices (see Table 2). Two types of point mutations were used: (i) the substitution of methionine to alanine; and, (ii) the substitution of methionine to isoleucine. Investigation of the TM5 peptide sequence revealed that two separate PxxxM motifs existed within the same helix, with a methionine residue present at M177^{5.38} as well the methionine residue identified at M193^{5.54}. Each of these methionine residues was mutated to alanine. We also mutated M193^{5.54} to isoleucine because a conserved PxxxI motif is found in the related family of P2Y receptors at the same location as the originally-identified PxxxM motif in A_{2A}R.

Specifically interacting residues in the TM5-M177A simulation set were identical to those identified in wild type TM5-TM5 dimers (Figure 7a) and included the M193^{5.54}xxxVY197^{5.58} motif. M177^{5.38} was not directly involved in the dimerisation between the two helices in any simulation. The specific interactions observed in the TM5- M193I^{5.54} simulation set were almost identical to those found in the wild-type but included I193^{5.54} in the interaction, despite the loss of the methionine at position 193 (Figure 7b). In contrast, the TM5- M193A^{5.54} mutation completely changed the contact interface of the helices (Figure 7c) and the key interacting residues were identified at a similar distance, but contained within a novel V196^{5.57}YxR199^{5.60}

motif. This provides a molecular explanation for the finding that mutation of the full-length A_{2A}R at position M193^{5,54} noticeably alters the monomer:dimer ratio, as observed with SDS-PAGE²⁷. Mutation of M193A^{5,54} causes a change in the way in which the two helices come together that prevents formation of TM5 homodimers, emphasizing the importance of the M193^{5,54} residue in the specificity of TM5-TM5 dimer formation *in vivo*.

3.3 Comparison with Experimental Structural Data

In order to assess the validity of our method, it is necessary to compare our results with experimental values. Our computational results closely match the experimental biophysical data of A_{2A} receptor homodimers and provide information about the nature of the contact interface between the two helices that cannot be determined experimentally. We then wished to see if we could obtain findings in agreement with experimentally-obtained structural data. Dimerization in Class A GPCRs involves the transmembrane domains, as opposed to Class C GPCRs, where dimerization is mediated by the large N terminal domain of the protein⁵⁹. We identified three additional dimeric Class A GPCRs in the PDB database that fulfilled the following criteria: 1) the crystallographic asymmetric unit is a dimer; 2) the software-determined (PISA) quaternary structure is a dimer; and, 3) the dimeric quaternary structure has been confirmed functionally. Rhodopsin, the CXCR4 chemokine receptor and the β_1 adrenergic receptor were chosen for further study; their corresponding TM helices (listed in Table 4) were constructed as described in section 2.3 and used in ensemble-based simulations.

3.3.1 Rhodopsin (1N3M)

Rhodopsin has been shown to exist in a native oligomeric form²⁰ and an atomic model of the rhodopsin dimer has been proposed as a working model for G protein-coupled receptors¹⁹. Three contact points between the rhodopsin monomers have been reported. The first is considered to be the strongest, with the largest contact area (578\AA^2) and is located between TM4 and TM5. The second exhibits a contact area of 333\AA^2 and is located between TM1 and TM2. The third contact point is considered the weakest interaction and is found between rows of dimers at the extracellular ends of TM1 with a contact area of 146\AA^2 ,^{19,22}. We ran two heterologous simulations between rhodopsin helices TM1 and TM2 and between rhodopsin helices TM4 and TM5 (see Table 4) to identify whether contact interfaces could be identified for either. Figure 8 shows that for both simulation sets, stable dimers were established, confirming that our computational method is able to produce results in agreement with structural data. In each case, the mean distance between helices was $\sim 7.6\text{-}8\text{\AA}$. The mean distance between specific interacting residues in the TM1-TM2 simulation (Figure 8a) is further apart than the mean distance between specific interacting residues in the TM4-TM5 simulation (Figure 8b).

3.3.2 CXCR4 (3ODU)

The crystal structure of the CXCR4 chemokine receptor bound to an antagonist small molecule IT1t has been reported and reveals a homodimer with an interface involving TM helices 5 and 6²³. We investigated interactions between helices in the CXCR4 receptor and identified the formation of stable dimers with specific interactions (Figure 9). We first ran a heterologous simulation between TM5 and TM6 and were unable to identify any interactions. CXCR4 is able to form homodimers in the absence of ligand⁶⁰ that are unable to be dissociated by a peptide derived from TM6⁶¹, suggesting that in unliganded CXCR4, the dimer interface may reside

between TM5 and TM5, in a manner analogous to the A_{2A} receptor. We then ran a CXCR4 TM5-TM5 simulation and identified, from the averaged interhelix contact matrices, the formation of dimers with specific interactions between F201^{5.40} and the following six residues: V198^{5.37}, Q200^{5.39}, F201^{5.40}, Q202^{5.41}, I204^{5.43} and M205^{5.44}.

3.3.3 β_1 -Adrenergic Receptor (4GPO)

Two alternating dimer interfaces have been proposed from the crystal structure of the ligand-free basal state of the β_1 adrenergic receptor (β_1 AR). The first involves TM1, TM2, extracellular loop 1 and the C-terminal H8; the second involves TM4 and TM5²⁴. We ran two heterologous simulations between β_1 AR helices TM1 and TM2 and between β_1 AR helices TM4 and TM5 (see Table 4) to identify whether contact interfaces could be identified for either. No stable dimers were formed in the TM1-TM2 simulation (Figure 10a). We investigated the possibility that the contact interface was formed between the two TM1 helices. We ran a TM1-TM1 simulation and identified a stable dimer in only one replica in the ensemble (Figure 10b). Stable dimers were formed between TM4 and TM5 with specific interactions identified between L159^{4.43} and Y231^{5.58} and between W166^{4.50} and Y277^{5.62} (Figure 10c).

3.4 Atomistic Representation and Proposed Nature of Interactions

In CG simulations, a small group of atoms is treated as a single particle (in a 4:1 ratio), a representation that lacks the specific details needed to describe the nature and type of the interactions that might take place when the two TM helices are within 10Å of each other. Representative atomistic structures were generated from CG models to enable a measurement of

distance between atoms⁴⁹ allowing hypotheses to be drawn about the molecular nature and possible role of the interactions between dimeric helices.

Figure 11 shows a representation of the converted atomistic wild-type A_{2A} TM5 dimer. Using this atomistic representation, the presence of possible electrostatic interactions or hydrogen bonding was investigated by measuring the distance between the specific interacting residues. The interaction between the two methionine residues (M193^{5.54}-M193^{5.54}) and between valine and tyrosine (V196^{5.57}-Y197^{5.58}) are likely to correspond to van der Waals interactions. Y197^{5.58} in helix 1 and Y197^{5.58} in helix 2 each interact as hydrogen donor and acceptor in the dimer, forming bonds between the peptide backbone and the tyrosine side chain (see Figure 11). As the measurement of these distances is longer than the optimal hydrogen bond distance, 2.7 Å, such hydrogen bonds are more likely to be formed backbone-to-side-chain because their interhelical distance of 8 Å is above the 7.6 Å limit of backbone-to-backbone interactions⁴⁷.

The rhodopsin dimer model (1N3M), shown in Figure 12a, reveals that there is a greater interface area between TM4 and TM5 than between TM1 and TM2. The specific interacting residues identified from the atomistic representation obtained using our computational method are distributed throughout the length of TM1 and TM2 but restricted to the bottom third of TM4 and TM5, with respect to the intracellular face of the receptor (Figure 12b-e). A comparison of our results with the TM1 and TM2 contact interface of 1N3M is shown in Figure 12b and c, respectively. The measured distance between the hydrogen on the COOH group of M119^{1.39} and the double-bonded oxygen of the COOH group on the side chain of D224^{2.50} is 10.32±3.21 Å in our model (Figure 12b), similar to 9.01 Å in 1N3M (Figure 12c). Measurement of the distance between F127^{1.47} and L218^{2.44} is 14.20±4.07 Å in our model and 15.28 Å in 1N3M. F127^{1.47} and L218^{2.44} are located towards the bottom of their respective helices, a position that is constrained

by the first intracellular loop of rhodopsin in 1N3M, but not in our model. Similar conservation of distance was identified between interacting residues in TM4 and TM5 (see Table 5).

Our studies of CXCR4 identified novel interactions in the homodimer between TM5 and TM5. This is similar to what was seen for A_{2A}, but the interacting residues in CXCR4 are closer to the extracellular side of the membrane than in A_{2A}. A comparison of the mean distance between interacting residues obtained from the simulations with the distance measured between the same residues in the crystal structure shows a similar conservation of distance, particularly between interacting residues further down the helix. This suggests a contribution of the loops in influencing interactions towards the ends of the helices, as was seen for rhodopsin.

Like rhodopsin, contact interfaces between TM1 and TM2 and between TM4 and TM5 had been proposed for the β_1 adrenergic receptor. However, using our method it is possible to identify a contact interface between TM1 and TM1, rather than between TM1 and TM2. Our measurements of distance are in agreement with those of the crystal structure. Our data suggest that the TM4-TM5 contact interface, and the four specific amino acids identified within it, may constitute the principal dimer interface in β_1 AR homodimers. It was not possible to compare the distances obtained in the TM4-TM5 simulation with those measured in the crystal structure 4GPO, which had been submitted with the orientation of the dimer showing the proposed TM1-TM2 interface.

4. CONCLUSIONS

In the present study, we have developed and assessed a method of ensemble-based coarse-grained classical molecular dynamics that we have used to predict protein-protein interactions

between TM helices of dimeric GPCRs. We applied our method to four different homomeric GPCRs for which experimental data exist and compare our predicted results with published experimental data. We have found that, in each case, the ensemble-based CG-MD methodology provides a reproducible measurement of the distance between interacting helices that corresponds well with the experimental data and is within the range of distances at which protein-protein interactions occur.

The first case was that of the A_{2A} adenosine receptor which had been shown experimentally to form homodimeric receptors through interactions between the TM5 helices of the two monomers. Our results identified specific interactions involving the PxxxM motif of TM5 and, specifically, at the M193^{5.54} residue within that motif. Our method accurately identified residues shown experimentally to be involved in TM5 homodimerisation. In parallel with work done experimentally, we investigated the role of M193^{5.54} by characterizing the M193^{5.54}A mutation. From this, we identified that the contact interface of the helices was completely changed and that the key interacting residues identified in the wild type conformation had moved to a new position, preventing formation of TM5 homodimers. Our results provide a molecular explanation for the experimental finding that the M193^{5.54}A mutation alters the monomer:dimer ratio at a level of detail that could not be determined biophysically and would require structural biology studies to confirm experimentally. The second case we examined was that of the rhodopsin dimer for which crystallographic data had identified contact interfaces between TM1 and TM2 and between TM4 and TM5. Ensemble CG-MD confirmed dimerization and the identification of specific interactions within each of these heterologous TM pairs. There is a striking convergence between the distances predicted computationally and those calculated from 1N3M, particularly for specific interactions between TMs 1 and 2, showing that our method is able to provide

accurate and precise predictions in agreement with experimental findings. Our method is also able to identify novel interfaces as seen in the third (CXCR4) and fourth (β_1 AR) cases we studied, where we identified a novel interface in CXCR4 between TM5 and TM5 and a novel interface in β_1 AR between TM1 and TM1, in addition to confirming the previously identified contact interface between TM4 and TM5 in the β_1 AR. The β_1 AR has been shown to form transient interactions whereas the β_2 adrenergic receptor can form stable oligomers⁶². Our ability to detect a stable dimer of TM1-TM1 in the β_1 AR shows the value of ensemble-based simulations for the identification of transient interactions.

We note that in the four cases we studied, there appears to be a pattern emerging of the nature and location of the contact interfaces. We observe either a single interface, at TM5 in both the A_{2A} and the CXCR4, or two contact interfaces, as seen in rhodopsin and the β_1 AR, one of which involves TM1 and the other between TM4 and TM5. Interestingly, interactions in TM5 are observed in both cases. As a greater number of dimeric GPCR crystal structures with corresponding biophysical and functional data become available, the conservation of the pattern we have detected should become clearer.

Our results unequivocally demonstrate that sufficient conformational sampling is required in coarse-grained MD to obtain reproducible and reliable results. In our simulations, we identified that several of the replicas within the ensemble failed to show any interactions and that a number of others began to interact late in the simulation at a point when accurate estimates of distance could no longer be achieved. A single trajectory simulation, particularly if either of these circumstances were to occur, would give inaccurate and potentially misleading results. Indeed, as we have repeatedly emphasized, ensembles are required to obtain accurate and precise results. We used error analysis to determine appropriate choices for ensemble size and run length. For

ensemble size, we observed that the rate of change in the standard deviation of the mean distance between helices decreased with increasing replica size and found that approximately 30 replicas was sufficient per ensemble to obtain reproducible results. For run length, we observed that the rate of increase in the standard deviation of the mean distance between helices increased with increasing run length, but that the rate of increase slowed substantially after approximately 300 ns. Interestingly, the negative control we included in our simulations showed no variation in the standard deviation of the mean distance between helices as a function of run length and a low standard deviation with a very rapid decrease to a constant value at an ensemble size of ~15 replicas. This behavior was notably different from simulations in which interactions were identified and provides a means of confirming the absence of interaction.

In conclusion, we have provided a systematic, reproducible and reliable protocol for determining the specific points of interaction between GPCR dimers. Our method discriminates between residues in TM helices that form specific interactions and residues that are in close proximity but do not interact. Our work extends the recent findings of ensemble-based fully atomistic MD studies, which have shown that an ensemble-based approach is required to generate predictions of protein properties that correlate well with experimental data. Our method is of great utility in further understanding GPCR function and also has broad applicability to many different types of membrane proteins, including receptor tyrosine kinases, ion channels, transporters and oligomeric complexes of various combinations of these.

FIGURES

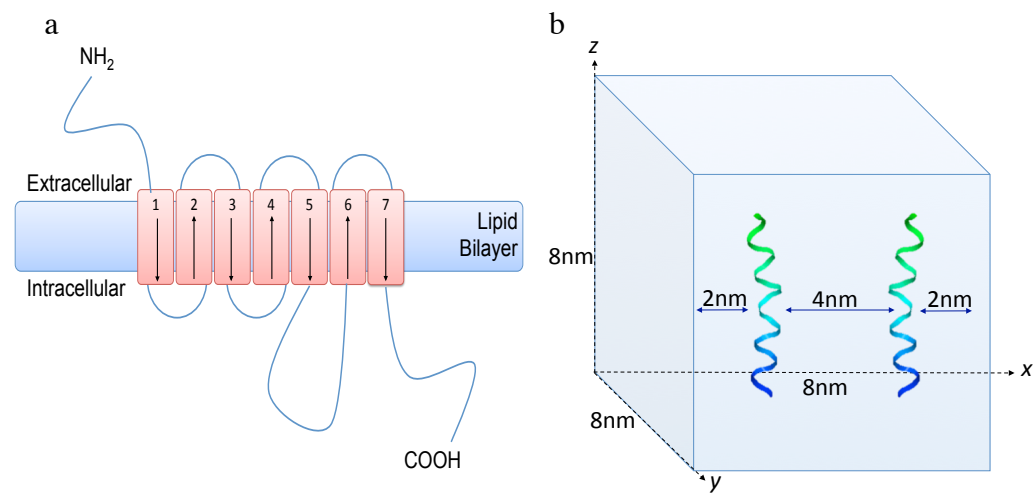


Figure 1. Experimental system showing (a) a schematic representation of the A_{2A} receptor structure indicating the directionality of the TM helices within the lipid bilayer and (b) placement of the TMs within the simulation box prior to addition of lipid and water.

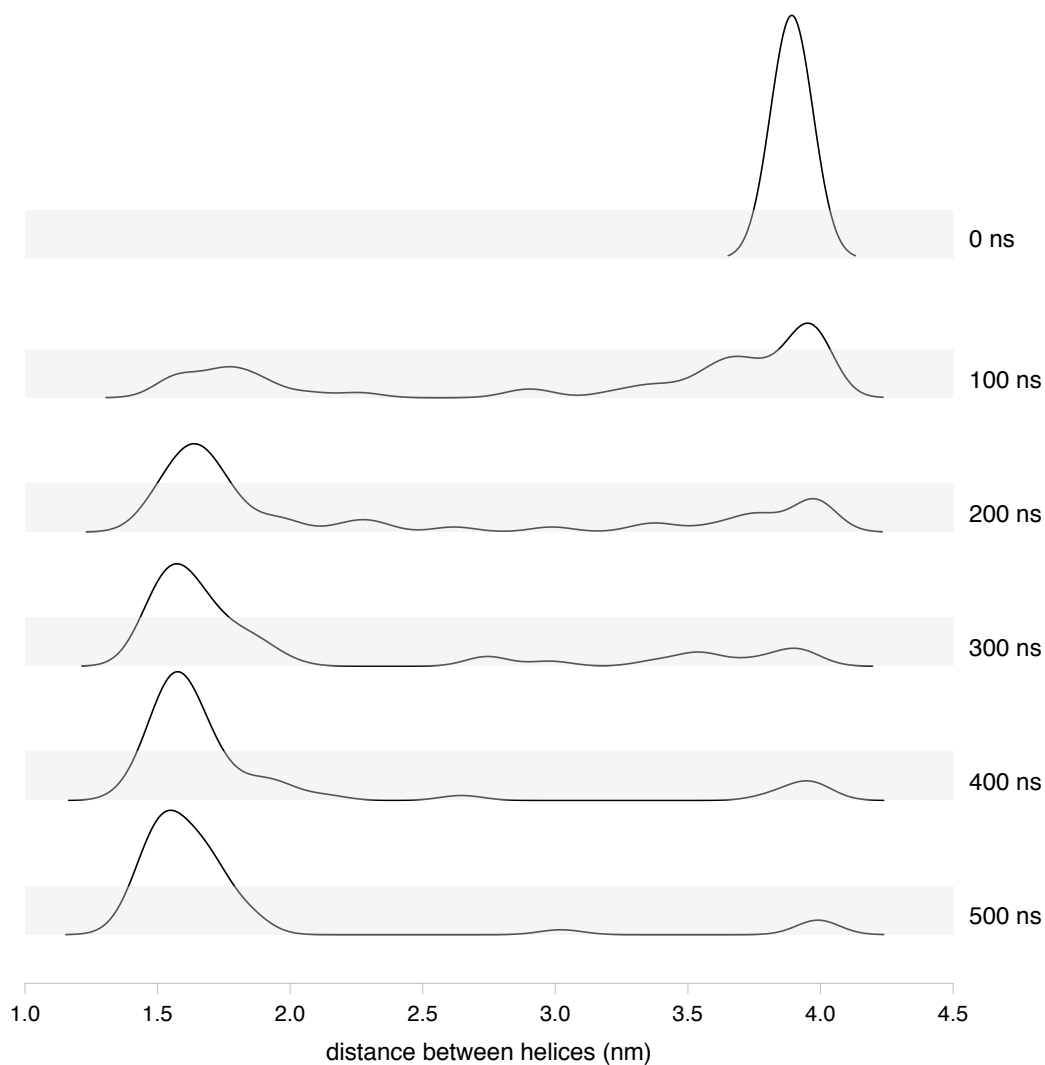


Figure 2. Distribution of the mean distance between the two TM5-TM5 wild type helices at 0, 100, 200, 300, 400 and 500 ns in all 50 replicas.

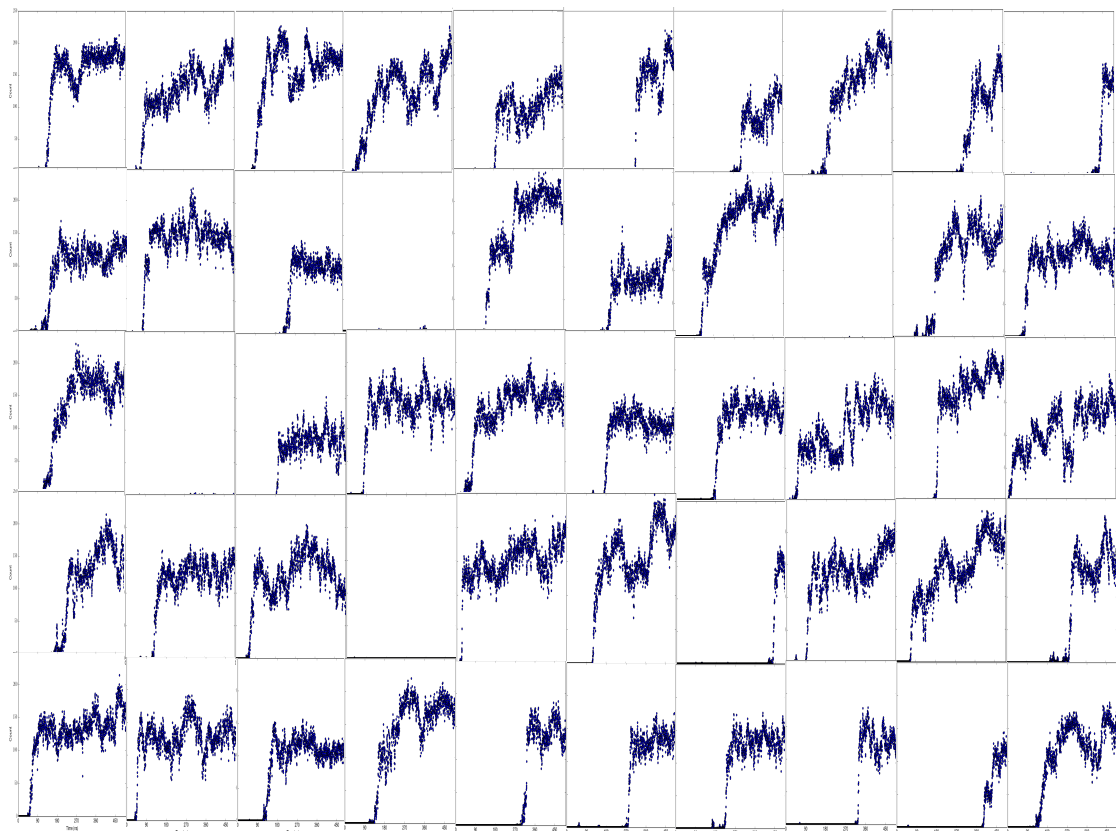


Figure 3. The number and timing of pairwise interactions for each of the 50 replicas within the wild-type TM5-TM5 dimer ensemble are shown. The x and y axes are linear and represent run length from 0 to 500ns and the number of interaction events from 0 to 250 counts, respectively.

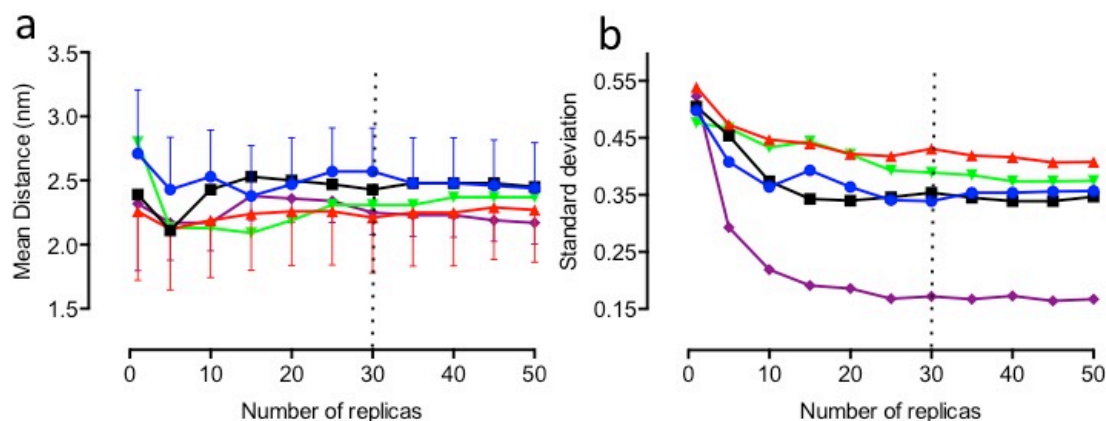


Figure 4. Variation in (a) the mean distance between TM helices and (b) the error (standard deviation) is shown as a function of the number of replicas performed for the following simulation sets: (●) wild-type TM5-TM5 helices, (■) M177A-mutated TM5-TM5 helices, (▲) M193A-mutated TM5-TM5 helices, (▼) M193I-mutated TM5-TM5 helices, and (◆) wild-type TM2-TM2 helices.

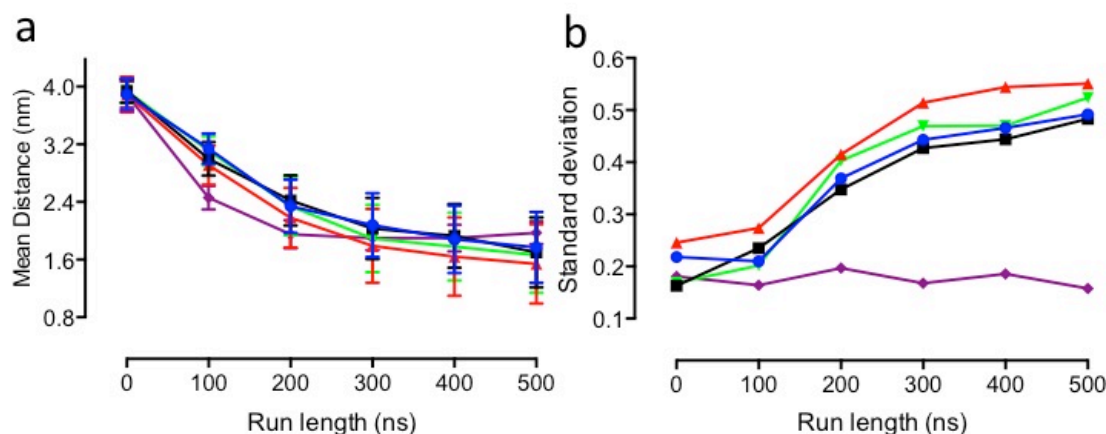


Figure 5. Variation in (a) the mean distance between TM helices and (b) the error (standard deviation) is shown as a function of the run length for the following simulation sets: (●) wild-type TM5-TM5 helices, (■) M177A-mutated TM5-TM5 helices, (▲) M193A-mutated TM5-TM5 helices, (▼) M193I-mutated TM5-TM5 helices, and (◆) wild-type TM2-TM2 helices.

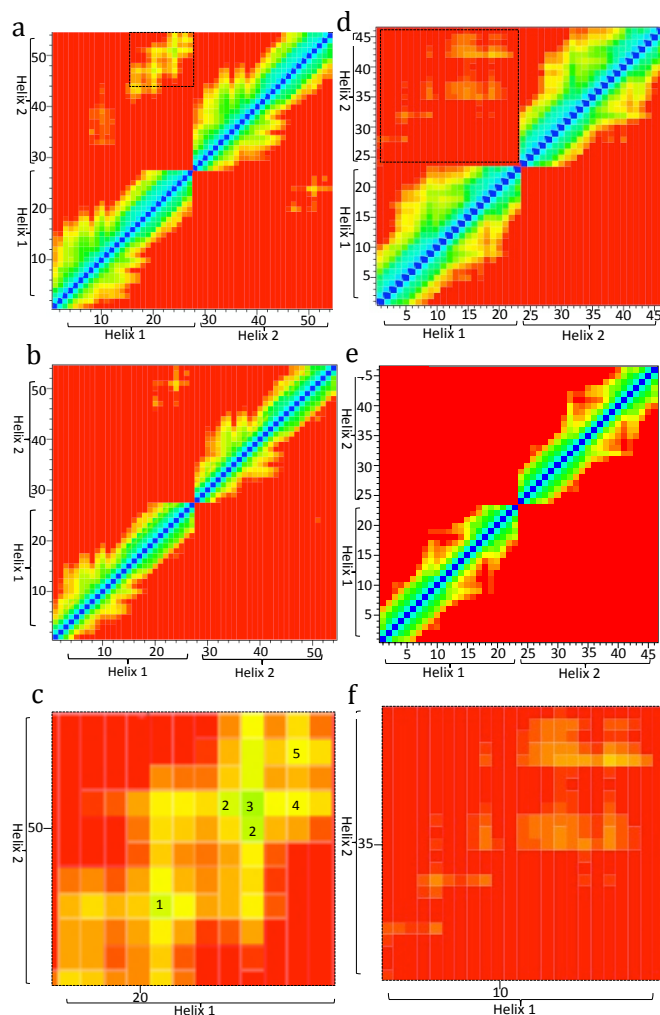


Figure 6. Contact matrices (heat maps) showing specific interactions between residues, as measured by distance, between two A_{2A} helices (“Helix 1” and “Helix 2”) for the wild-type TM5-TM5 simulation (a-c) and for the TM2-TM2 negative control (d-f). Results shown are the average for each ensemble. Interhelical distances at the 15 Å cutoff are shown in the top left quarter of panels (a) and (d). Interhelical distances at the 12 Å cutoff are shown in the top left corner of (b) and (e) and in the lower right quarter of panels (a) and (d). Interhelical distances at the 10 Å cutoff are shown in the lower right quarter of panels (b) and (e). The region shown in the black rectangle in (a) and (d) is magnified in (c) and (f), respectively. The five numbered interactions shown in (c) are identified in Table 3. The color scale indicates distance between helices: blue corresponds to 0 Å (superposition of the two helical backbones at all cutoffs); green corresponds to 5 Å (10 Å cutoff), 6 Å (12 Å cutoff), 7.5 Å (15 Å cutoff); yellow corresponds to 7 Å (10 Å cutoff), 8 Å (12 Å cutoff), 12 Å (15 Å cutoff); red corresponds to the cutoff distances applied (10 Å, 12 Å or 15 Å).

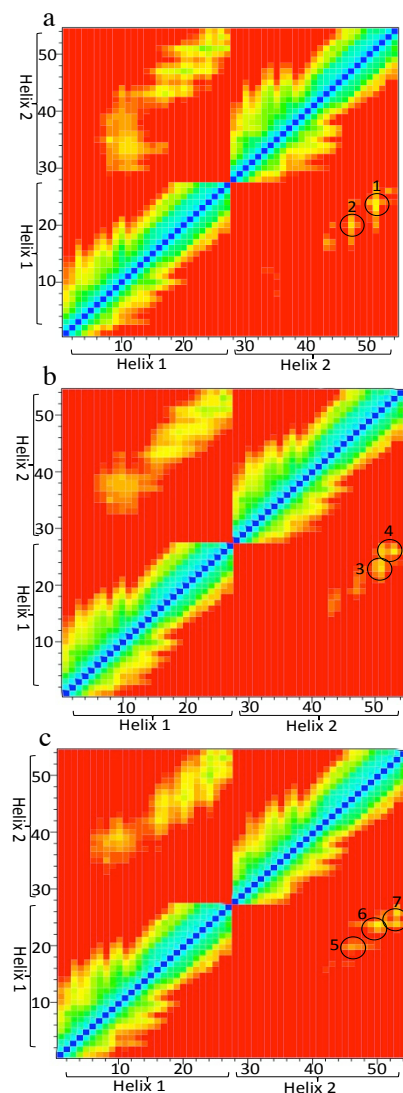


Figure 7. Contact matrices (heat maps) showing specific interactions between two mutated A_{2A} TM5 helices (“Helix 1” and “Helix 2”) with the following residues mutated: M177A (a), M193^{5.54}A (b) and M193^{5.54}I (c). Results shown are the average for each ensemble. Interhelical distances at the 15Å and 12Å cutoff distances are shown in the top left quarter and in the lower right quarter of panels (a-c), respectively. The color scale is as indicated in Figure 6. Circles indicate areas with key interhelical contacts. The identified amino acid interactions are numbered as follows: (1) M193^{5.54} with M193^{5.54}; (2, 3) V196^{5.57} with Y197^{5.58} and Y197^{5.58} with Y197^{5.58}; (4) Y197^{5.58} with I200^{5.61} and R199^{5.60} with R199^{5.60}; (5) L192^{5.53} with I193^{5.54}, V196^{5.57} with Y197^{5.58} and Y197^{5.58} with R199^{5.60}, (6) Y197^{5.58} with I200^{5.61} and Y197^{5.58} with R199^{5.60}; and (7) R199^{5.60} with R199^{5.60}.

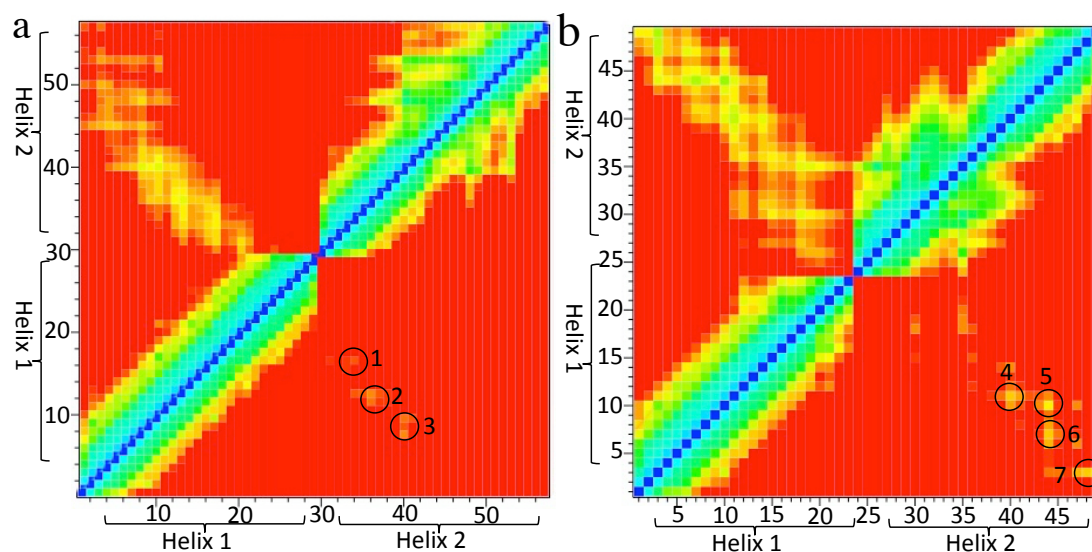


Figure 8. Contact matrices (heat maps) between two rhodopsin helices, showing specific interactions between TM1 (Helix 1) and TM2 (Helix 2) (a) and between TM4 (Helix 1) and TM5 (Helix 2) (b). Results shown are the average for each ensemble. The color scale is as indicated in Figure 6. Circles indicate areas with key interhelical contacts. The identified amino acid interactions are numbered as follows: 1) F127^{1.47} with L218^{2.44}; 2) L122^{1.42} with L220^{2.46}, I123^{1.43} with L218^{2.44} and N219^{2.45} with L220^{2.46}; 3) Y118^{1.38} with D224^{2.50}, M119^{1.39} with D224^{2.50}, F120^{1.40} with D224^{2.50} and F120^{1.40} with L225^{2.51}; 4) F418^{4.48} with L521^{5.51} and T419^{4.49} with L521^{5.51}; 5) F418^{4.48} with F525^{5.55}; 6) M414^{4.44} with F525^{5.55}, G415^{4.45} with F525^{5.55} and V416^{4.46} with F525^{5.55}; 7) H411^{4.41} with Y528^{5.58}, H411^{4.41} with G529^{5.59} and H411^{4.41} with Q530^{5.60}.

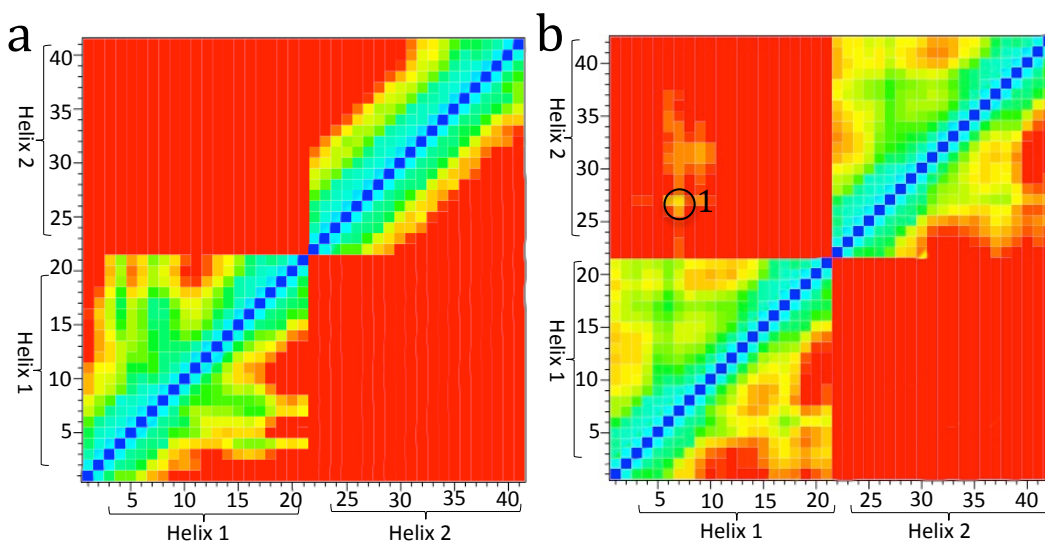


Figure 9. Contact matrices (heat maps) between two CXCR4 helices, showing specific interactions (a) between TM5 (Helix 1) and TM6 (Helix 2) and (b) between TM5 (Helix 1) and TM5 (Helix 2). The identified amino acid interactions are numbered as follows: 1) F201^{5.40} with V198^{5.37}, Q200^{5.39}, F201^{5.40}, Q202^{5.41}, I204^{5.43} and M205^{5.44}. Table 5 shows a comparison of the distances between specific atoms in interacting residues of the representative structures and the distances between the same atoms in the model structure.

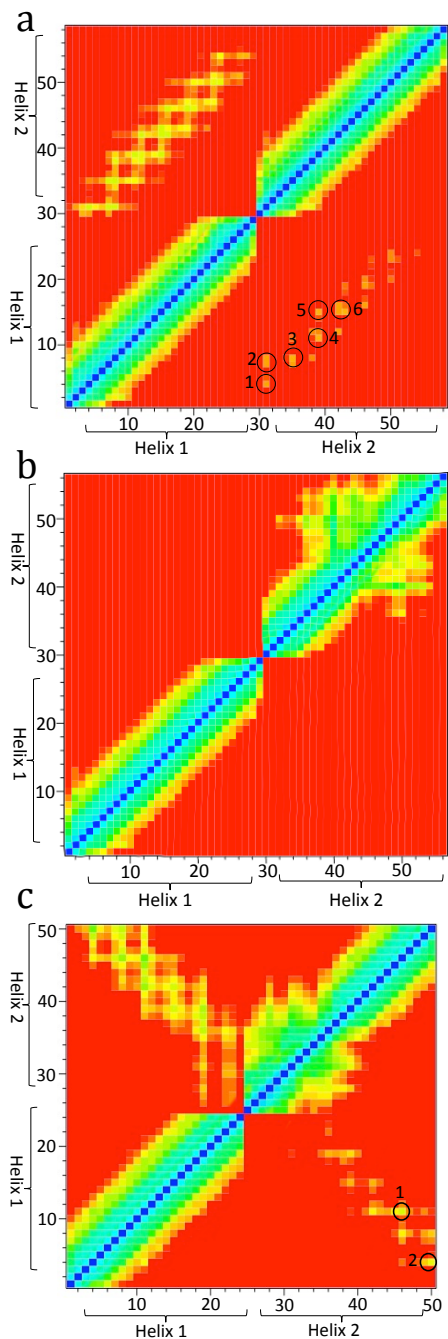


Figure 10. Contact matrices (heat maps) between two β_1 AR helices, showing specific interactions (a) between TM1 (Helix 1) and TM2 (Helix 2), (b) between TM1 (Helix 1) and TM1 (Helix 2) and (c) between TM4 (Helix 1) and TM5 (Helix 2) (c). Results shown are the average for each ensemble. The color scale is as indicated in Figure 6. Circles indicate areas with key interhelical contacts. The identified amino acid interactions are numbered as follows: in (a) 1) W40^{1.31} with A42^{1.33}, S45^{1.36} and L46^{1.37}; 2) M44^{1.35} with L46^{1.37}; 3) A49^{1.39} with M48^{1.38}; 4) L53^{1.44} with M48^{1.38}; 5) L53^{1.44} with M48^{1.38}, V51^{1.40} and V52^{1.41}; 6) L54^{1.45} with V51^{1.40}; (b) 1) L159^{4.43} with Y231^{5.58}; 2) W166^{4.50} with Y227^{5.62}. Table 5 shows a comparison of the distances between specific atoms in interacting residues of the representative structures and the distances between the same atoms in the model structure.

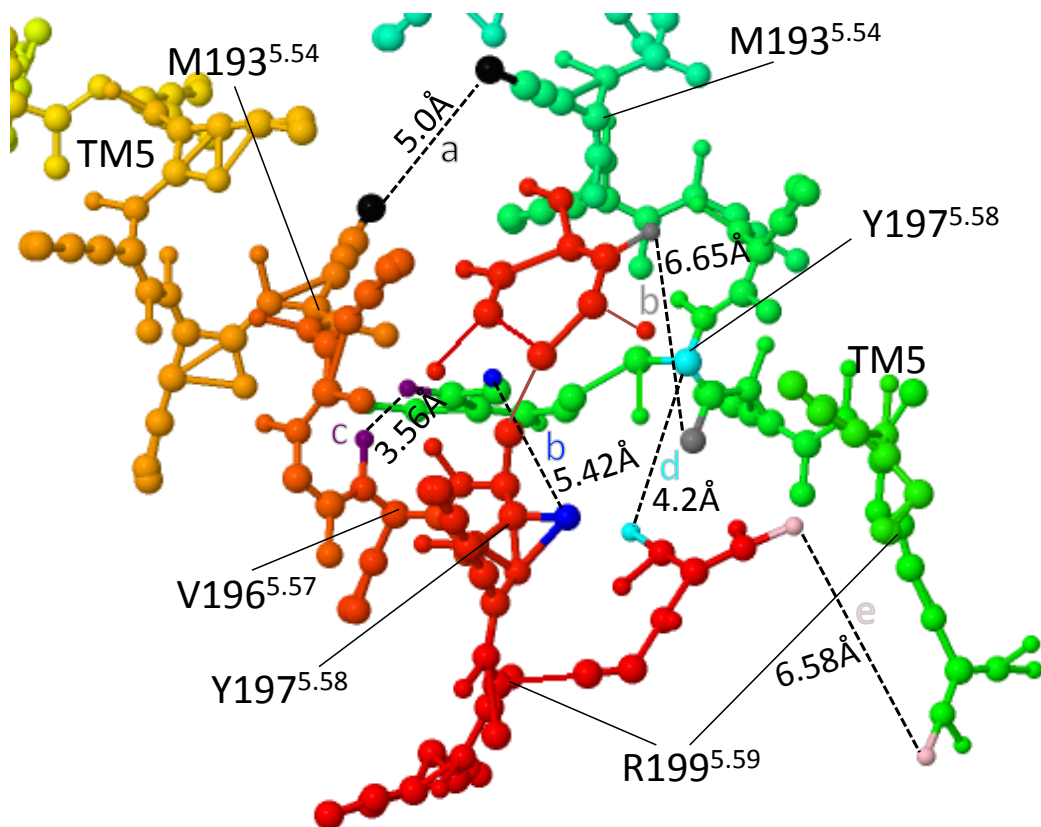


Figure 11. Atomistic representation of the pairwise interactions identified from the wild-type TM5-TM5 ensemble. The representative mean distance is shown in the figure and the mean distance \pm SD for all hits detected per pair is shown in Table 3. All distances between interacting amino acids are calculated from side chain to side chain.

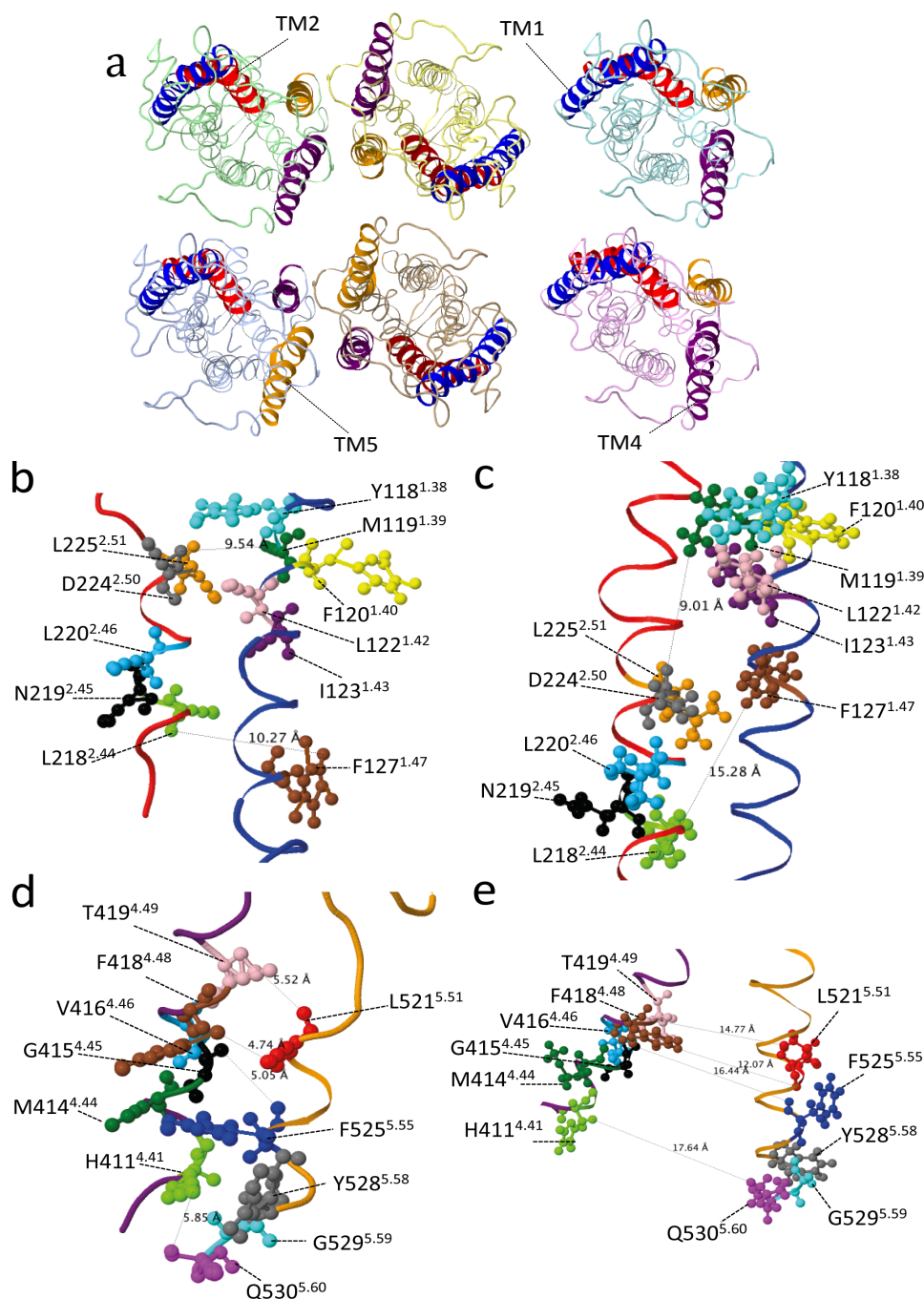


Figure 12. (a) Atomistic structure of the rhodopsin dimer model (1N3M) viewed from above, with the TMs used for simulations identified by color as follows: TM1 (blue), TM2 (red), TM4 (purple) and TM5 (orange). Representative TM structures were obtained from the means of all replicas in which interactions were detected. The representative and model structures of TM1-TM2 are shown in (b) and (c), respectively. The representative and model structures of TM4-TM5 are shown in (d) and (e), respectively. Specific interactions were identified in TM1-TM2 simulations (M119^{1.39} with D224^{2.50} and F127^{1.47} with L218^{2.44}) and in TM4-TM5 simulations (H411^{4.41} with Q530^{5.60}, G415^{4.45} with F525^{5.55}, F418^{4.48} with L521^{5.51} and T419^{4.49} with L521^{5.51}). Table 5 shows a comparison of the distances between specific atoms in interacting residues of the representative structures and the distances between the same atoms in the model structure.

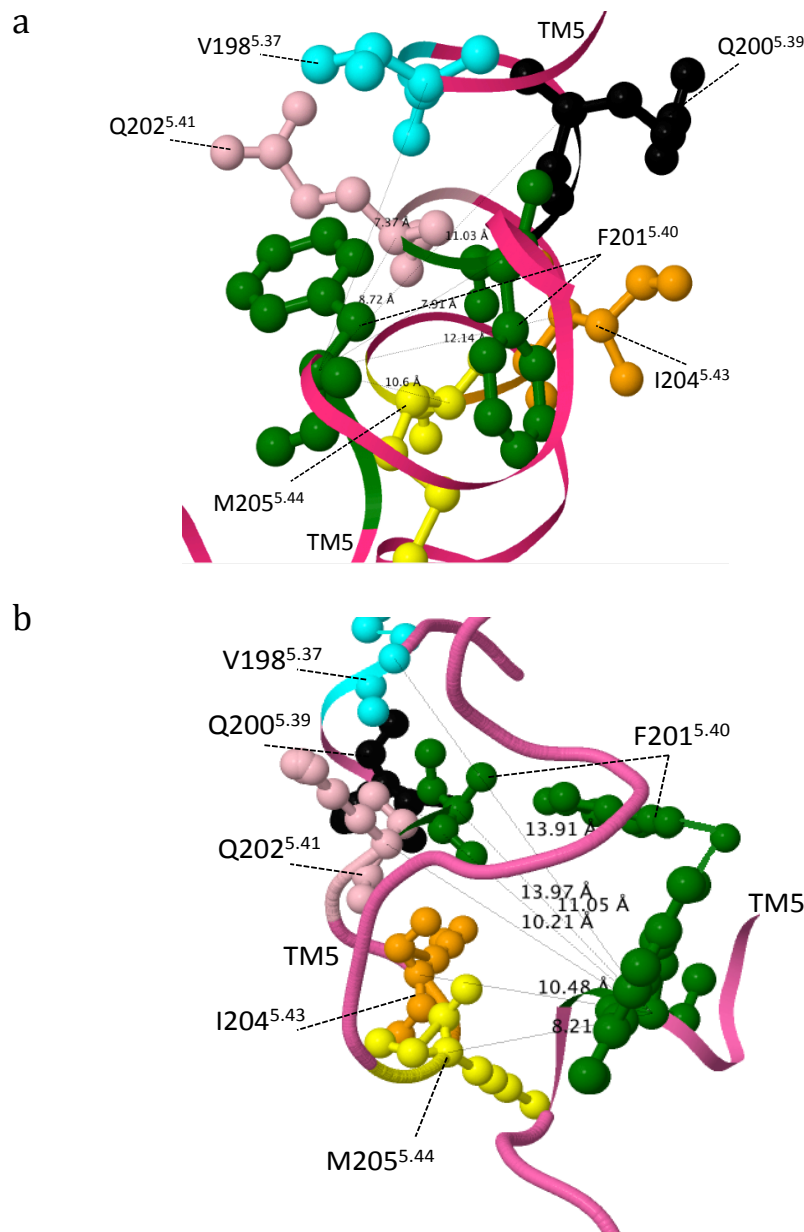


Figure 13. (a) Atomistic structure of the CXCR4 dimer model (3ODU), with the TMs used for simulations identified by color where is TM5 (pink). Representative TM structures were obtained from the means of all replicas in which interactions were detected. The representative and model structures of TM5-TM5 are shown in (a) and (b), respectively. Specific interactions were identified in TM5-TM5 simulations (F201^{5.40} with V198^{5.37}, Q200^{5.39}, F201^{5.40}, Q202^{5.41}, I204^{5.43} and M205^{5.44}). Table 5 shows a comparison of the distances between specific atoms in interacting residues of the representative structures and the distances between the same atoms in the model structure.

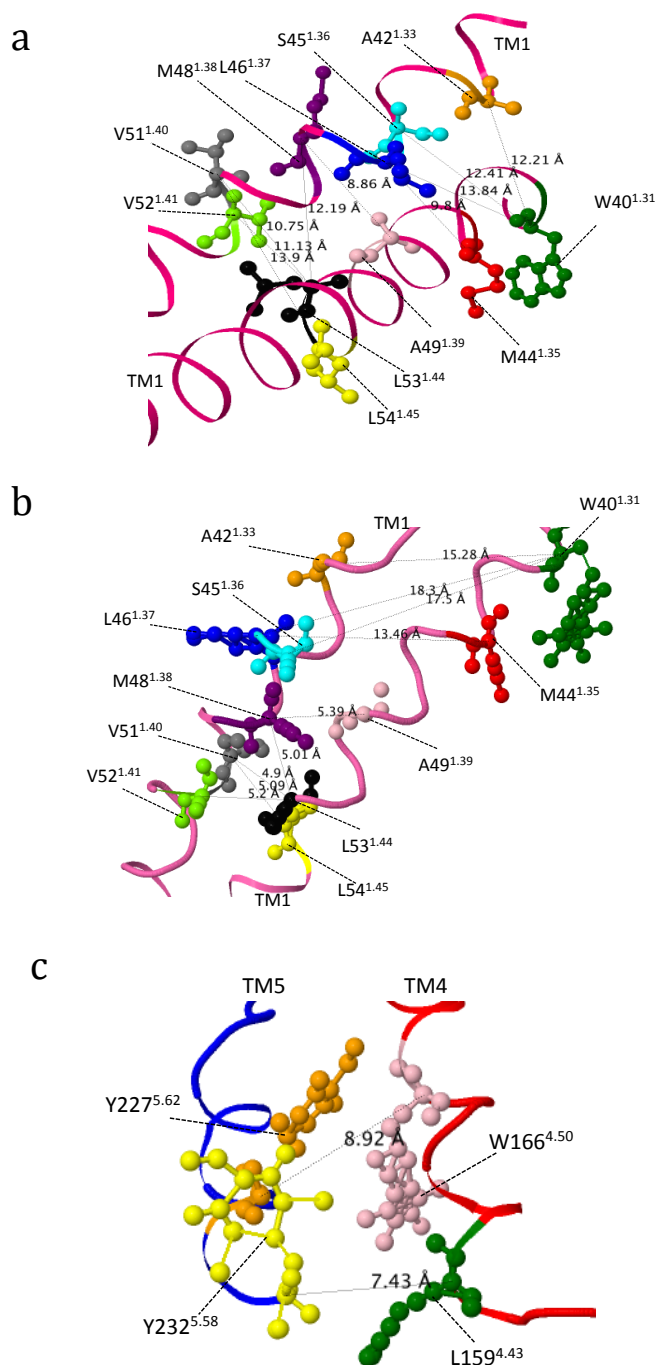


Figure 14. (a) Atomistic structure of the β_1 AR dimer model (4GPO), with the TMs used for simulations identified by color as follows: TM1 (pink), TM4 (red) and TM5 (blue). Representative TM structures were obtained from the means of all replicas in which interactions were detected. The representative and model structures of TM1-TM1 are shown in (a) and (b) respectively. The representative and model structures of TM4-TM5 are shown in (c). Specific interactions were identified in TM1-TM1 simulations

(W40^{1.31} with A42^{1.33}, S45^{1.36} and L46^{1.37}; M44^{1.35} with L46^{1.37}; A49^{1.39} with M48^{1.38}; L53^{1.44} with M48^{1.38}; L53^{1.44} with M48^{1.38}, V51^{1.40} and V52^{1.41}; L54^{1.45} with V51^{1.40}) and in TM4-TM5 simulations (L159^{4.43} with Y231^{5.58}; W166^{4.50} with Y227^{5.62}). Table 5 shows a comparison of the distances between specific atoms in interacting residues of the representative structures and the distances between the same atoms in the model structure.

TABLES

Table 1. Computational Methods Used for Modeling Mammalian GPCR Dimers

| Type | GPCR dimers | Method | Force field | Interface | Number of replicas | Time scale | Reference |
|--------------|--|-------------|-------------------------------------|--|--|--------------------|--|
| Homodimers | Rho/Rho | Docking | CVFF | TM4,5/TM4,5 | | 100 ps | Filipek <i>et al.</i> , 2004 ⁶³ |
| | | | | TM1,2/TM1,2 | | ND* | Han <i>et al.</i> , 2009 ⁶⁴ |
| | | MD | Gromos-87 OPLSAA Amber/parm99 | TM4,5/TM4,5 | 1 | 45 ns | Filizola <i>et al.</i> , 2006 ⁶⁶ |
| | | | | | 1 | 0.1 μ s | Cordomi and Perez, 2009 ⁶⁷ |
| | | CG-MD | Martini | TM1,2, H8 | 2 | 300 ns | Neri <i>et al.</i> , 2010 ⁶⁸ |
| | | | | | 1 ^a | 8 μ s | Periole <i>et al.</i> , 2007 ³⁶ |
| | β_2 -/ β_2 -adrenergic | CG-MD | Martini | TM4,5 TM6,7 | 10 | 100 μ s | Periole <i>et al.</i> , 2012 ⁶⁹ |
| | | | | | 2 | 5 μ s | Ghosh <i>et al.</i> , 2014 ⁴¹ |
| | | | | H8/H8 | 4 | \sim 200 μ s | Prasanna <i>et al.</i> , 2014 ⁴⁰ |
| | | | | | 1 ^b | \sim 18 μ s | Mondal <i>et al.</i> , 2013 ³⁹ |
| | β_1 -/ β_1 -adrenergic | CG-MD | Martini | TM1/TM1 TM5/TM5 | 3 ^c (2 runs different starting point) | 2 μ s | Mondal <i>et al.</i> , 2013 ³⁹ |
| | | | | | | | |
| | α_{1B} -/ α_{1B} -adrenergic | Docking** | | TM5/TM5 TM6/TM6 TM7/TM7 | | ND | Fanelli <i>et al.</i> , 1999 ⁷⁰ |
| | | | | | | | |
| | 5-HT ₄ /5-HT ₄ | Docking | | TM2,4/TM2,4 TM4,6/TM4,6 | | ND | Soulier <i>et al.</i> , 2005 ⁷¹ |
| | | | | | | | Russo <i>et al.</i> , 2007 ⁷² |
| | 5-HT _{1A} /5-HT _{1A} | Docking | CHARMM | TM4,5/TM4,5 | | 15 ns | Berthouze <i>et al.</i> , 2007 ⁷³ |
| | | | | | | | Gorinski <i>et al.</i> , 2012 ⁷⁴ |
| | CXCR4/CXCR4 | Docking | | TM4,5,IL2/TM4,5 | | ND | Kaczor <i>et al.</i> , 2013 ⁶⁵ |
| | | | | | | | Rodriguez <i>et al.</i> , 2012 ⁷⁵ |
| | NTSR ₁ /NTSR ₁ | MD*** | OPLSAA | TM3/TM4,5 TM5/TM5 | 1 | 50 ns | |
| | | | | | | | |
| | NTSR ₁ /NTSR ₁ | Docking | CHARMM | TM1/TM4 TM4/TM4 | | ND | Casciari <i>et al.</i> , 2008 ⁷⁶ |
| | | | | | | | |
| | δ -OR/ δ -OR | CG-MD | Martini | TM2,3,4/TM2,3,4 TM4/TM4 favored over TM4,5/TM4,5 | 1 ^d | 250 ns | Provasi <i>et al.</i> , 2010 ³⁷ |
| | | | | | 2 | 1.5 μ s | Johnston <i>et al.</i> , 2011 ⁷⁷ |
| | κ -OR/ κ -OR | Docking | | TM1/TM2 | | ND | Kaczor <i>et al.</i> , 2013 ⁶⁵ |
| | A _{2A} R/A _{2A} R | Docking | CHARMM | TM1,2,3/TM1,2,3 TM1/TM1 TM1,4/TM1,4 TM2,3/TM2,3 TM6,7/TM6,7 H8,13/TM6 | | ND | Fanelli and Felling, 2011 ⁷⁸ |
| | | | | | | | |
| | A ₃ R/A ₃ R | MD | Amber7 FF99 | TM4,5/TM4,5 | 1 | 500 ps | Kim and Jacobson, 2006 ⁷⁹ |
| | TXA ₂ /TXA ₂ | Docking | | TM1/TM1 TM1/TM2,EL2 H8/H8 | | ND | Fanelli <i>et al.</i> , 2011 ⁸⁰ |
| | | | | | | | |
| | D ₂ R/D ₂ R | Monte Carlo | | ND | | ND | Woolf and Linderman, 2004 ⁸¹ |
| | SSTR ₁ /SSTR ₁ | Monte Carlo | | ND | | ND | Woolf and Linderman, 2004 ⁸¹ |
| | LHR-LHR | Docking | | TM4/TM6,7 TM4/TM4 | | ND | Fanelli 2007 ⁸⁰ |
| | | | | | 1 | 1 ns | Fanelli 2007 ⁸⁰ |
| Heterodimers | A _{2A} R/D ₂ R | Docking | | TM3,4/TM5,6 TM3,4,5/TM4,5 TM4,5/TM3,4,5 | | ND | Canals <i>et al.</i> , 2003 ²⁵ |
| | | | | | | | |
| | mGluR ₂ /5-HT _{2A} | Docking | | TM4,5/TM4,5 | | ND | Bruno <i>et al.</i> , 2009 ³³ |
| | | | | | 1 | 40 ns | Bruno <i>et al.</i> , 2009 ³³ |
| | μ -OR/ δ -OR | Docking | CHARMM22/27 | TM6,7/TM4,5 TM1,7/TM4,7 | | ND | Liu <i>et al.</i> , 2009 ⁸² |
| | | | | | 1 | 5 ns | Liu <i>et al.</i> , 2009 ⁸² |
| Homotetramer | (V ₂ R) ₄ | MD | CHARMM22/27 | TM3,4/TM4,7 TM4,5/TM4,5 | | 5 ns | Witt <i>et al.</i> , 2007 ³² |
| | | | | | 1 | | |

* ND: Not defined, IL: intracellular loop, EL: extracellular loop; ^a 4 different structures; ⁵ 9 different structures; ^c 2 runs, different starting point; ^d umbrella sampling of 43 different starting points

Table 2. Sequences of the A_{2A}R helices used in ensemble simulation sets.

| A _{2A} R Helices | Sequences ^a |
|-----------------------------|--|
| TM2-Wild-type | FVVS L AAAD52 ^{2..50} I A VG V LAIPFAITI |
| TM5-Wild-type | MNYM VYFNFFACVLV P189 ^{5..50} LLL M LG V YLRI |
| TM5-M177 ^{5..38} A | MNY <u>A</u> VYFNFFACVLV P189 ^{5..50} LLL M LG V YLRI |
| TM5-M193 ^{5..54} A | MNYM VYFNFFACVLV P189 ^{5..50} LLL <u>A</u> LG V YLRI |
| TM5-M193 ^{5..54} I | MNYM VYFNFFACVLV P189 ^{5..50} LLL <u>I</u> LG V YLRI |

^a Residues suggested to play a key role in the dimerization of the A_{2A}R TM5 helix²⁷ are indicated in bold; mutated residues are underlined and italicized in (blue). The conserved amino acid for each TM helix is shown in red and is numbered both using the Ballesteros and Weinstein nomenclature (superscript) and by residue number.

Table 3. The number of interactions (hits) for specific interacting residues identified in the contact matrices for the wild-type TM5-TM5 simulation at the 10Å cut-off.

| Figure Label | Interacting residues | Replica number | | | | | | | | | | | Mean Distance ± Standard Deviation (in Å) |
|-----------------|--|----------------|---|----|----|----|----|----|----|----|----|----|--|
| | | 1 | 5 | 10 | 15 | 20 | 25 | 30 | 35 | 40 | 45 | 50 | |
| 1 | M193 ^{5.54} -M193 ^{5.54} | 0 | 0 | 1 | 1 | 1 | 2 | 2 | 3 | 5 | 6 | 6 | 7.59 ± 2.89 |
| 2 | V196 ^{5.57} -Y197 ^{5.58} | 0 | 1 | 2 | 2 | 2 | 4 | 4 | 6 | 9 | 12 | 13 | 9.16 ± 2.5 |
| 3 | Y197 ^{5.58} -Y197 ^{5.58} | 0 | 1 | 1 | 1 | 1 | 1 | 2 | 4 | 6 | 7 | 8 | 9.11 ± 2.85 |
| 4 | Y197 ^{5.58} -R199 ^{5.60} | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 3 | 5 | 6 | 6 | 9.83 ± 3.57 |
| 5 | R199 ^{5.60} -R199 ^{5.60} | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 3 | 4 | 5 | 5 | 8.06 ± 2.99 |

Table 4. Sequences of the rhodopsin, CXCR4 and β_1 AR receptor helices used in ensemble simulation sets.

| Receptor | Helices | Sequences ^a |
|--------------|---------|--|
| Rhodopsin | TM1 | QFSMLAAYMFLIMLGFPIN ^{1.50} (55)FLTLYVTVQ |
| | TM2 | NYILLNLAVAD ^{2.50} (83)LFMVFGGFTTTLYTSLH |
| | TM4 | ENHAIMGVAFTW ^{4.50} (161)VMALACAAPPL |
| | TM5 | NESFVIYMFVVHFII ^{5.50} (215)LIVIFFCYGQ |
| CXCR4 | TM5 | VVVFQFQHIMVGLIL ^{5.50} (211)GIVIL |
| | TM6 | VILILAFFACWL ^{6.50} (254)YYIGISI |
| β_1 AR | TM1 | QWEAGMSLLMALVVLLIVAGN ^{1.50} (59)VLVIAAIG |
| | TM2 | NLFITSLACAD ^{2.50} (87)LVMGLLVVPFGATLVV |
| | TM4 | ARAKVIICTVW ^{4.50} (166)AISALVSFLPIMM |
| | TM5 | AYAIASSIISFYI ^{5.50} (219)LLIMIFVYLRVY |

^a The conserved amino acid for each TM helix is shown in red and is numbered using both the Ballesteros and Weinstein nomenclature (superscript) and by residue number.

Table 5. The comparison of distance of the identified interacting residues[†] from contact matrix graphs of the rhodopsin, CXCR4 and β_1 AR helices and the crystal rhodopsin dimer (1N3M), CXCR4 dimer (4GPO) and the β_1 AR dimer (3ODU)

| Receptor | Helices | Interacting residues | Crystal structure distance (Å) | Mean Distance (Å) ± Standard deviation |
|--------------|---------|---|--------------------------------|--|
| Rhodopsin | TM1-TM2 | F127 ^{1.47} - L218 ^{2.44} | 15.28 | 14.2 ± 4.07 |
| | | M119 ^{1.39} - D224 ^{2.50} | 9.01 | 10.32 ± 3.1 |
| | TM4-TM5 | F418 ^{4.48} - L521 ^{5.51} | 12.07 | 9.18 ± 3.24 |
| | | T419 ^{4.49} - L521 ^{5.51} | 14.77 | 8.13 ± 1.96 |
| | | G415 ^{4.45} - F525 ^{5.55} | 16.44 | 7.5 ± 2.3 |
| | | H411 ^{4.41} - Q530 ^{5.60} | 17.64 | 9.06 ± 3.34 |
| CXCR4 | TM5-TM5 | F201 ^{5.40} - V198 ^{5.37} | 7.37 | 15.55 ± 3.34 |
| | | F201 ^{5.40} - Q200 ^{5.39} | 11.03 | 13.7 ± 1.46 |
| | | F201 ^{5.40} - F201 ^{5.40} | 7.91 | 13.6 ± 2.89 |
| | | F201 ^{5.40} - Q202 ^{5.41} | 8.72 | 14.93 ± 3.13 |
| | | F201 ^{5.40} - I204 ^{5.43} | 12.14 | 13.11 ± 3.37 |
| | | F201 ^{5.40} - M205 ^{5.44} | 10.6 | 12.6 ± 4.96 |
| β_1 AR | TM1-TM1 | W40 ^{1.31} - A42 ^{1.33} | 12.21 | 15.28* |
| | | W40 ^{1.31} - S45 ^{1.36} | 12.41 | 17.5* |
| | | W40 ^{1.31} - L46 ^{1.37} | 13.84 | 18.3* |
| | | M44 ^{1.35} - L46 ^{1.37} | 9.8 | 13.46* |
| | | A49 ^{1.39} - M48 ^{1.38} | 8.86 | 5.39* |
| | | L53 ^{1.44} - M48 ^{1.38} | 12.19 | 5.01* |
| | | L53 ^{1.44} - V51 ^{1.40} | 10.75 | 4.9* |
| | | L53 ^{1.44} - V52 ^{1.41} | 11.13 | 5.2* |
| | | L54 ^{1.45} - V51 ^{1.40} | 13.9 | 5.09* |
| | TM4-TM5 | L159 ^{4.43} - Y231 ^{5.58} | ND** | 9.01 ± 2.22 |
| | | W166 ^{4.50} - Y227 ^{5.62} | ND** | 7.9 ± 1.99 |

[†] Distances are measured from backbone to backbone.
* Interactions were detected in only one replica in the ensemble.
** Not Determined (ND): The distances between TM4 and TM5 could not be measured due to the orientation of the dimer in the 4GPO crystal structure, which is submitted showing the TM1-TM2 dimer interface.

AUTHOR INFORMATION

Corresponding Author

*E-mail: a.townsend-nicholson@ucl.ac.uk

Author Contributions

The manuscript was written through contributions of all authors. All authors have given approval to the final version of the manuscript.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENT

The authors would like to acknowledge the support of EU H2020 CompBioMed project (<http://www.compbioimed.eu/>, 675451), the MRC Medical Bioinformatics project (MR/L016311/1), and funding from the UCL Provost. N Altwaijry was supported by a King Saud University Studentship. The authors acknowledge the use of the UCL Legion and Grace High Performance Computing Facilities (Legion@UCL, Grace@UCL) and associated support services, in the completion of this work.

REFERENCES

1. Sadiq, S. K.; Wright, D. W.; Kenway, O. A.; Coveney, P. V., Accurate ensemble molecular dynamics binding free energy ranking of multidrug-resistant HIV-1 proteases. *J. Chem. Inf. Model* **2010**, 50 (5), 890.
2. Wan, S.; Coveney, P. V., Rapid and accurate ranking of binding affinities of epidermal growth factor receptor sequences with selected lung cancer drugs. *J. R. Soc. Interface* **2011**, 8 (61), 1114.
3. Wright, D. W.; Hall, B. A.; Kenway, O. A.; Jha, S.; Coveney, P. V., Computing Clinically Relevant Binding Free Energies of HIV-1 Protease Inhibitors. *J. Chem. Theory Comput.* **2014**, 10, 1228.

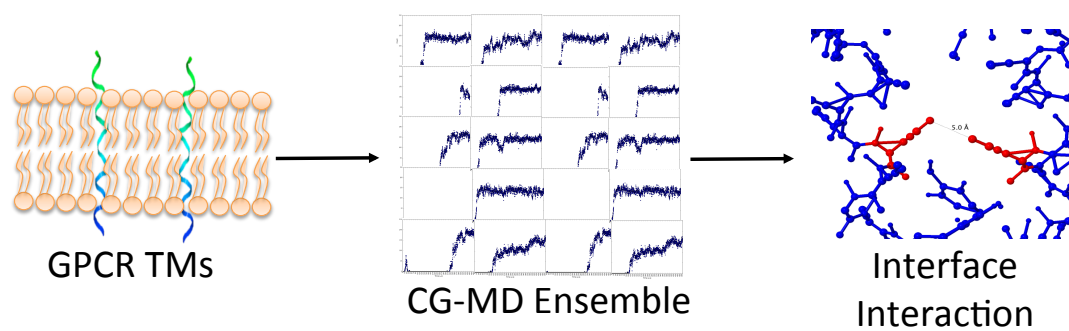
4. Wan, S.; Knapp, B.; Wright, D. W.; Deane, C. M.; Coveney, P. V., Rapid, Precise, and Reproducible Prediction of Peptide-MHC Binding Affinities from Molecular Dynamics That Correlate Well with Experiment. *J. Chem. Theory. Comput.* **2015**, *11* (7), 3346.
5. Coveney, P. V.; Wan, S., On the calculation of equilibrium thermodynamic properties from molecular dynamics. *Phys. Chem. Chem. Phys.* **2016**, *18*, 30236.
6. Bhati, A. P.; Wan, S. Z.; Wright, D. W.; Coveney, P. V., Rapid, Accurate, Precise, and Reliable Relative Free Energy Prediction Using Ensemble Based Thermodynamic Integration. *J. Chem. Theory Comput.* **2017**, *13* (1), 210.
7. Wan, S.; Bhati, A. P.; Zasada, S. J.; Wall, I.; Green, D.; Bamborough, P.; Coveney, P. V., Rapid and Reliable Binding Affinity Prediction of Bromodomain Inhibitors: a Computational Study. *J. Chem. Theory Comput.* **2017**, *3*, 784.
8. King, B. F.; Townsend-Nicholson, A., Involvement of P2Y1 and P2Y11 purinoceptors in parasympathetic inhibition of colonic smooth muscle. *J. Pharmacol. Exp. Ther.* **2008**, *324* (3), 1055.
9. Devader, C.; Drew, C. M.; Geach, T. J.; Tabler, J.; Townsend-Nicholson, A.; Dale, L., A novel nucleotide receptor in *Xenopus* activates the cAMP second messenger pathway. *FEBS Lett.* **2007**, *581* (27), 5332.
10. Dalpiaz, A.; Townsend-Nicholson, A.; Beukers, M. W.; Schofield, P. R.; AP, I. J., Thermodynamics of full agonist, partial agonist, and antagonist binding to wild-type and mutant adenosine A1 receptors. *Biochem. Pharmacol.* **1998**, *56* (11), 1437.
11. Atkinson, M. R.; Townsend-Nicholson, A.; Nicholl, J. K.; Sutherland, G. R.; Schofield, P. R., Cloning, characterisation and chromosomal assignment of the human adenosine A3 receptor (ADORA3) gene. *Neurosci. Res.* **1997**, *29* (1), 73.
12. Townsend-Nicholson, A.; Schofield, P. R., A threonine residue in the seventh transmembrane domain of the human A1 adenosine receptor mediates specific agonist binding. *J. Biol. Chem.* **1994**, *269* (4), 2373.
13. Townsend-Nicholson, A.; Shine, J., Molecular cloning and characterisation of a human brain A1 adenosine receptor cDNA. *Brain Res Mol Brain Res* **1992**, *16* (3-4), 365.
14. Selbie, L. A.; Townsend-Nicholson, A.; Iismaa, T. P.; Shine, J., Novel G protein-coupled receptors: a gene family of putative human olfactory receptor sequences. *Brain Res. Mol. Brain Res.* **1992**, *13* (1-2), 159.
15. Shonberg, J.; Kling, R. C.; Gmeiner, P.; Löber, S., GPCR crystal structures: Medicinal chemistry in the pocket. *Bioorg. Med. Chem.* **2015**, *23*, 3880.
16. George, S. R.; O'Dowd, B. F.; Lee, S. P., G-protein-coupled receptor oligomerization and its potential for drug discovery. *Nat. Rev. Drug Discovery* **2002**, *1* (10), 808.
17. Jones, K. A.; Borowsky, B.; Tamm, J. A.; Craig, D. A.; Durkin, M. M.; Dai, M.; Yao, W. J.; Johnson, M.; Gunwaldsen, C.; Huang, L. Y.; Tang, C.; Shen, Q. R.; Salon, J. A.; Morse, K.; Laz, T.; Smith, K. E.; Nagarathnam, D.; Noble, S. A.; Branchek, T. A.; Gerald, C., GABA(B) receptors function as a heteromeric assembly of the subunits GABA(B)R1 and GABA(B)R2. *Nature* **1998**, *396* (6712), 674.
18. Jordan, B. A.; Devi, L. A., G-protein-coupled receptor heterodimerization modulates receptor function. *Nature* **1999**, *399* (6737), 697.
19. Liang, Y.; Fotiadis, D.; Filipek, S.; Saperstein, D. A.; Palczewski, K.; Engel, A., Organization of the G protein-coupled receptors rhodopsin and opsin in native membranes. *J. Biol. Chem.* **2003**, *278* (24), 21655.

20. Fotiadis, D.; Liang, Y.; Filipek, S.; Saperstein, D. A.; Engel, A.; Palczewski, K., Atomic-force microscopy: Rhodopsin dimers in native disc membranes. *Nature* **2003**, *421* (6919), 127.
21. Bai, M., Dimerization of G-protein-coupled receptors: roles in signal transduction. *Cell. Signalling* **2004**, *16*, 175.
22. Fotiadis, D.; Jastrzebska, B.; Philippsen, A.; Muller, D. J.; Palczewski, K.; Engel, A., Structure of the rhodopsin dimer: a working model for G-protein-coupled receptors. *Curr. Opin. Struct. Biol.* **2006**, *16* (2), 252.
23. Wu, B.; Chien, E. Y.; Mol, C. D.; Fenalti, G.; Liu, W.; Katritch, V.; Abagyan, R.; Brooun, A.; Wells, P.; Bi, F. C.; Hamel, D. J.; Kuhn, P.; Handel, T. M.; Cherezov, V.; Stevens, R. C., Structures of the CXCR4 chemokine GPCR with small-molecule and cyclic peptide antagonists. *Science* **2010**, *330* (6007), 1066.
24. Huang, J.; Chen, S.; Zhang, J. J.; Huang, X. Y., Crystal structure of oligomeric beta1-adrenergic G protein-coupled receptors in ligand-free basal state. *Nat. Struct. Mol. Biol.* **2013**, *20* (4), 419.
25. Canals, M.; Marcellino, D.; Fanelli, F.; Ciruela, F.; de Benedetti, P.; Goldberg, S. R.; Neve, K.; Fuxe, K.; Agnati, L. F.; Woods, A. S.; Ferre, S.; Lluís, C.; Bouvier, M.; Franco, R., Adenosine A2A-dopamine D2 receptor-receptor heteromerization: qualitative and quantitative assessment by fluorescence and bioluminescence energy transfer. *J. Biol. Chem.* **2003**, *278* (47), 46741.
26. Canals, M.; Burgueno, J.; Marcellino, D.; Cabello, N.; Canela, E. I.; Mallol, J.; Agnati, L.; Ferre, S.; Bouvier, M.; Fuxe, K.; Ciruela, F.; Lluís, C.; Franco, R., Homodimerization of adenosine A2A receptors: qualitative and quantitative assessment by fluorescence and bioluminescence energy transfer. *J. Neurochem.* **2004**, *88* (3), 726.
27. Thevenin, D.; Lazarova, T.; Roberts, M. F.; Robinson, C. R., Oligomerization of the fifth transmembrane domain from the adenosine A2A receptor. *Protein Sci.* **2005**, *14* (8), 2177.
28. Thevenin, D.; Lazarova, T., Stable interactions between the transmembrane domains of the adenosine A2A receptor. *Protein Sci.* **2008**, *17*, 1188.
29. Le, F.; Townsend-Nicholson, A.; Baker, E.; Sutherland, G. R.; Schofield, P. R., Characterization and chromosomal localization of the human A2a adenosine receptor gene: ADORA2A. *Biochem. Biophys. Res. Commun.* **1996**, *223* (2), 461.
30. Filizola, M.; Weinstein, H., The study of G-protein coupled receptor oligomerization with computational modeling and bioinformatics. *FEBS J.* **2005**, *272*, 2926.
31. Simpson, L. M.; Taddese, B.; Wall, I. D.; Reynolds, C. A., Bioinformatics and molecular modelling approaches to GPCR oligomerization. *Curr. Opin. Pharmacol.* **2010**, *10* (1), 30.
32. Witt, M.; Ślusarz, M. J.; Ciarkowski, J., Molecular Modeling of Vasopressin V2 Receptor Tetramer in Hydrated Lipid Membrane. *QSAR Comb. Sci.* **2008**, *27* (6), 684.
33. Bruno, A.; Entrena Guadix, A. E.; Costantino, G., Molecular Dynamics Simulation of the Heterodimeric mGluR2/5HT2A Complex. An Atomistic Resolution Study of a Potential New Target in Psychiatric Conditions. *J. Chem. Inf. Model.* **2009**, *49*, 1602.
34. Vanni, S.; Neri, M.; Tavernelli, I.; Rothlisberger, U., Observation of “Ionic Lock” Formation in Molecular Dynamics Simulations of Wild-Type $\beta 1$ and $\beta 2$ Adrenergic Receptors. *Biochem.* **2009**, *48*, 4789.
35. Rodríguez, D.; Ranganathan, A.; Carlsson, J., Strategies for Improved Modeling of GPCR-Drug Complexes: Blind Predictions of Serotonin Receptors Bound to Ergotamine. *J. Chem. Inf. Model.* **2014**, *54*, 2004.

36. Periole, X.; Huber, T.; Marrink, S. J.; Sakma, T. P., G Protein-Coupled Receptors Self-Assemble in Dynamics Simulations of Model Bilayers. *J. Am. Chem. Soc.* **2007**, *129*, 10126.
37. Provasi, D.; Johnston, J. M.; Filizola, M., Lessons from Free Energy Simulations of δ - Opioid Receptor Homodimers Involving the Fourth Transmembrane Helix. *Biochem.* **2010**, *49*, 6771.
38. Periole, X.; Marrink, S. J., The Martini coarse-grained force field. *Methods Mol. Biol.* **2013**, *924*, 533.
39. Mondal, S.; Johnston, J. M.; Wang, H.; Khelashvili, G.; Filizola, M.; Weinstein, H., Membrane Driven Spatial Organization of GPCRs. *Sci. Rep.* **2013**, *3*, 2909.
40. Prasanna, X.; Chattopadhyay, A.; Sengupta, D., Cholesterol Modulates the Dimer Interface of the β_2 -Adrenergic Receptor via Cholesterol Occupancy Sites. *Bioph. J. Vol.* **2014**, *106*, 1290.
41. Ghosh, A.; Sonavane, U.; Joshi, R., Mutliscala modelling to understand the self-assembly mechanism of human β_2 -adrenergic receptor in lipid bilayer. *Comput. Bio. & Chem.* **2014**, *48*, 29.
42. Marrink, S. J.; Risselada, H. J.; Yefimov, S.; Tieleman, D. P.; de Vries, A. H., The MARTINI force field: coarse grained model for biomolecular simulations. *J. Phys. Chem. B* **2007**, *111* (27), 7812.
43. Monticelli, L.; Kandasamy, S. K.; Periole, X.; Larson, R. G.; Tieleman, D. P.; Marrink, S. J., The MARTINI coarse-grained force field: Extension to proteins. *J. Chem. Theory Comput.* **2008**, *4* (5), 819.
44. Psachoulia, E.; Fowler, P. W.; Bond, P. J.; Sansom, M. S., Helix-helix interactions in membrane proteins: coarse-grained simulations of glycophorin a helix dimerization. *Biochem.* **2008**, *47* (40), 10503.
45. Prakash, A.; Janosi, L.; Doxastakis, M., Self-association of models of transmembrane domains of ErbB receptors in a lipid bilayer. *Biophys. J.* **2010**, *99* (11), 3657.
46. Sahoo, D.; Weers, P. M.; Ryan, R. O.; Narayanaswami, V., Lipid-triggered conformational switch of apolipophorin III helix bundle to an extended helix organization. *J. Mol. Biol.* **2002**, *321* (2), 201.
47. Senes, A.; Ubarretxena-Belandia, I.; Engelman, D. M., The C-H...O hydrogen bond: A determinant of stability and specificity in transmembrane helix interactions. *Proc. Natl. Acad. Sci.* **2001**, *98*, 9056.
48. De Jong, D. H.; Singh, G.; Bennett, W. F. D.; Arnarez, C.; Wassenaar, T. A.; Schafer, L. V.; Periole, X.; Tieleman, D. P.; Marrink, S. J., Improved Parameters for the Martini Coarse-Grained Protein Force Field. *J. Chem. Theory Comput.* **2012**, *9* (1), 687.
49. Wassenaar, T. A.; Pluhackova, K.; Böckmann, R. A.; Marrink, S. J.; Tieleman, D. P., Going Backward: A Flexible Geometric Approach to Reverse Transformation from Coarse Grained to Atomistic Models. *J. Chem. Theory Comput.* **2014**, *10*, 676.
50. Daura, X.; Gademann, K.; Jaun, B.; Seebach, D.; van Gunsteren, W. F.; Mark, A. E., Peptide folding: When simulation meets experiment. *Angew. Chem. Int. Ed.* **1999**, *38* (1-2), 236.
51. Humphrey, W.; Dalke, A.; Schulten, K., VMD: visual molecular dynamics. *J Mol Graph* **1996**, *14* (1), 33-8, 27.
52. A., B. J.; H., W., Integrated Methods for the Construction of Three-Dimensional Models and Computational Probing of Structure-Function Relations in G Protein-Coupled Receptors. *Methods Neurosci.* **1995**, *25*.

53. Sali, A.; Potterton, L.; Yuan, F.; van Vlijmen, H.; Karplus, M., Evaluation of comparative protein modeling by MODELLER. *Proteins* **1995**, 23 (3), 318.
54. Fiser, A.; Sali, A., Modeller: generation and refinement of homology-based protein structure models. *Methods Enzymol.* **2003**, 374, 461.
55. Jaakola, V. P.; Griffith, M. T.; Hanson, M. A.; Cherezov, V.; Chien, E. Y.; Lane, J. R.; Ijzerman, A. P.; Stevens, R. C., The 2.6 angstrom crystal structure of a human A2A adenosine receptor bound to an antagonist. *Science* **2008**, 322 (5905), 1211.
56. Wassenaar, T. A.; Ingólfsson, H.; Böckmann, R. A.; Tieleman, P.; Marrink, S. J., Computational Lipidomics with insane: A Versatile Tool for Generating Custom Membranes for Molecular Simulations. *J. Chem. Theory Comput.* **2015**, 11, 2144.
57. Thevenin, D.; Roberts, M. F.; Lazarova, T.; Robinson, C. R., Identifying interactions between transmembrane helices from the adenosine A2A receptor. *Biochem.* **2005**, 44, 16239.
58. Vidia, P. A.; Chenb, J.; Irudayarajb, J. M. K.; Watts, V. J., Adenosine A2A receptors assemble into higher-order oligomers at the plasma membrane. *FEBS Lett.* **2008**, 582, 3985.
59. Levitz, J.; Habrian, C.; Bharill, S.; Fu, Z.; Vafabakhsh, R.; Isacoff, E. Y., Mechanism of Assembly and Cooperativity of Homomeric and Heteromeric Metabotropic Glutamate Receptors. *Neuron* **2016**, 92 (1), 143.
60. Babcock, G. J.; Farzan, M.; Sodroski, J., Ligand-independent dimerization of CXCR4, a principal HIV-1 coreceptor. *J. Biol. Chem.* **2003**, 278 (5), 3378.
61. Percherancier, Y.; Berchiche, Y. A.; Slight, I.; Volkmer-Engert, R.; Tamamura, H.; Fujii, N.; Bouvier, M.; Heveker, N., Bioluminescence resonance energy transfer reveals ligand-induced conformational changes in CXCR4 homo- and heterodimers. *J. Biol. Chem.* **2005**, 280 (11), 9895.
62. Dorsch, S.; Klotz, K. N.; Engelhardt, S.; Lohse, M. J.; Bunemann, M., Analysis of receptor oligomerization by FRAP microscopy. *Nat. Methods* **2009**, 6 (3), 225.
63. Filipek, S.; Krzysko, K. A.; Fotiadis, D.; Liang, Y.; Saperstein, D. A.; Engel, A.; Palczewski, K., A concept for G protein activation by G protein-coupled receptor dimers: the transducin/rhodopsin interface. *Photochem. Photobiol. Sci.* **2004**, 3 (6), 628.
64. Han, Y.; Moreira, I. S.; Urizar, E.; Weinstein, H.; Javitch, J. A., Allosteric communication between protomers of dopamine class A GPCR dimers modulates activation. *Nat. Chem. Biol.* **2009**, 5 (9), 688.
65. Kaczor, A. A.; Selent, J.; Sanz, F.; Pastor, M., Modeling complexes of transmembrane proteins: systematic analysis of protein-protein docking tools. *Mol. Inf.* **2013**, 32, 717.
66. Filizola, M.; Wang, S. X.; Weinstein, H., Dynamic models of G-protein coupled receptor dimers: indications of asymmetry in the rhodopsin dimer from molecular dynamics simulations in a POPC bilayer. *J. Comput.-Aided Mol. Des.* **2006**, 20 (7-8), 405.
67. Cordomi, A.; Perez, J. J., Structural rearrangements of rhodopsin subunits in a dimer complex: a molecular dynamics simulation study. *J. Biomol. Struct. Dyn.* **2009**, 27 (2), 127.
68. Neri, M.; Vanni, S.; Tavernelli, I.; Rothlisberger, U., Role of aggregation in rhodopsin signal transduction. *Biochem.* **2010**, 49 (23), 4827.
69. Periole, X.; Knepp, A. M.; Sakmar, T. P.; Marrink, S. J.; Huber, T., Structural determinants of the supramolecular organization of G protein-coupled receptors in bilayers. *J. Am. Chem. Soc.* **2012**, 134 (26), 10959.
70. Fanelli, F.; Menziani, C.; Scheer, A.; Cotecchia, S.; De Benedetti, P. G., Theoretical study on receptor-G protein recognition: New insights into the mechanism of the alpha 1b-adrenergic receptor activation. *Int. J. of Quantum Chem.* **1999**, 73 (2), 71.

71. Soulier, J. L.; Russo, O.; Giner, M.; Rivail, L.; Berthouze, M.; Onger, S.; Maigret, B.; Fischmeister, R.; Lezoualc'h, F.; Sicsic, S.; Berque-Bestel, I., Design and synthesis of specific probes for human 5-HT₄ receptor dimerization studies. *J. Med. Chem.* **2005**, *48* (20), 6220.
72. Russo, O.; Berthouze, M.; Giner, M.; Soulier, J. L.; Rivail, L.; Sicsic, S.; Lezoualc'h, F.; Jockers, R.; Berque-Bestel, I., Synthesis of specific bivalent probes that functionally interact with 5-HT₄ receptor dimers. *J. Med. Chem.* **2007**, *50* (18), 4482.
73. Berthouze, M.; Rivail, L.; Lucas, A.; Ayoub, M. A.; Russo, O.; Sicsic, S.; Fischmeister, R.; Berque-Bestel, I.; Jockers, R.; Lezoualc'h, F., Two transmembrane Cys residues are involved in 5-HT₄ receptor dimerization. *Biochem. Biophys. Res. Commun.* **2007**, *356* (3), 642.
74. Gorinski, N.; Kowalsman, N.; Renner, U.; Wirth, A.; Reinartz, M. T.; Seifert, R.; Zeug, A.; Ponimaskin, E.; Niv, M. Y., Computational and Experimental Analysis of the Transmembrane Domain 4/5 Dimerization Interface of the Serotonin 5-HT_{1A} Receptor. *Mol. Pharmacol.* **2012**, *82* (3), 448.
75. Rodriguez, D.; Gutierrez-de-Teran, H., Characterization of the homodimerization interface and functional hotspots of the CXCR4 chemokine receptor. *Proteins* **2012**, *80* (8), 1919.
76. Casciari, D.; Dell'Orco, D.; Fanelli, F., Homodimerization of neurotensin 1 receptor involves helices 1, 2, and 4: insights from quaternary structure predictions and dimerization free energy estimations. *J. Chem. Inf. Model.* **2008**, *48* (8), 1669.
77. Johnston, J. M.; Aburi, M.; Provati, D.; Bortolato, A.; Urizar, E.; Lambert, N. A.; Javitch, J. A.; Filizola, M., Making structural sense of dimerization interfaces of delta opioid receptor homodimers. *Biochem.* **2011**, *50* (10), 1682.
78. Fanelli, F.; Feline, A., Dimerization and ligand binding affect the structure network of A_{2A} adenosine receptor. *Biochim. Biophys. Acta* **2011**, 1256.
79. Kim, S. K.; Jacobson, K. A., Computational prediction of homodimerization of the A₃ adenosine receptor. *J. Mol. Graph. Modell.* **2006**, *25* (4), 549.
80. Fanelli, F., Dimerization of the lutropin receptor: insights from computational modeling. *Mol. Cell. Endocrinol.* **2007**, 260-262, 59.
81. Woolf, P. J.; Linderman, J. J., An algebra of dimerization and its implications for G-protein coupled receptor signaling. *J. Theor. Biol.* **2004**, 229, 157.
82. Liu, X.; Ming Kai, M.; Jin, J.; Wang, R., Computational study of the heterodimerization between l and d receptors. *J. Comput.-Aided Mol. Des.* **2009**, *23*, 321-332.



TOC Graphic